VOLUME XXVI MAY, 1934 PARTS I AND II

BIOMETRIKA

ON THE P_{λ_n} TEST FOR RANDOMNESS: REMARKS, FURTHER ILLUSTRATION, AND TABLE OF P_{λ_n} FOR GIVEN VALUES OF $-\log_{10}\lambda_n$.

BY FLORENCE N. DAVID.

DR KARL PEARSON has drawn my attention to the fact that in his recent paper on the P_{λ_n} Test for Randomness in Sampling*, he had judged roughly in which direction the probability integrals should be measured, endeavouring to choose that direction which would lead to the more stringent result, i.e. that in which the probability of a more improbable result occurring would be the smaller. Consider the simple case of an individual drawn at random from a normal parent population. If in the ordinary notation we found $\frac{1}{2}(1-\alpha) = \cdot 005$, we should argue that such an individual was very unlikely to have been drawn in a single random sampling, for it deviates from the mean value by about 2·58 times the standard deviation in excess, but actually the improbability would be as great in the case of an individual deviating from the mean value by 2·58 in defect, in which case the probability integral *measured in the same direction* would be ·995. Now, if we have two probability integrals p_1 and p_2, we may either take $p_1 \times p_2$ or $(1-p_1)(1-p_2)$ but we cannot take $p_1(1-p_2)$ or $(1-p_1)p_2$, for the proof of the randomness in this test depends on the probability integrals being measured in the *same* direction. Accordingly in this case $p_1 p_2$ will be larger than $(1-p_1)(1-p_2)$ if $p_1 + p_2$ be greater than unity. But when we increase the number of probability integrals there is no easy method of at once determining whether

$$p_1 p_2 \ldots p_n \quad \text{or} \quad (1-p_1)(1-p_2)\ldots(1-p_n)$$

will give the more stringent test. These remarks apply equally well when the frequency distribution of the parental population, true or hypothesised, is skew.

In order to ascertain whether any strikingly different results would arise if the probability integrals in Pearson's illustrations had been measured in the opposite sense, I have worked out the values of $-\log_{10}(1-p_1)(1-p_2)\ldots(1-p_n)$ for those illustrations and determined the corresponding values of $Q_{\lambda_n} = 1 - P_{\lambda_n}$. This seemed practically as short as comparing the two products

$$p_1 p_2 \ldots p_n \quad \text{and} \quad (1-p_1)(1-p_2)\ldots(1-p_n),$$

for it is actually the value of $I(n-1, u)$ we require for relative appreciation.

* *Biometrika*, Vol. xxv. pp. 379—410.

The following Table I gives the value of Q_{λ_n} for the various illustrations, and denotes which method leads to the more stringent results: Q_{λ_n} measures the probability of less probable cases occurring. The smaller it is, the less stringent the test.

TABLE I.

Illustration	Table and Page, *Biometrika*, Vol. xxv.	A Value of Q_{λ_n} from $p_1 p_2 \ldots p_n$	B Value of Q_{λ_n} from $(1-p_1)(1-p_2)\ldots(1-p_n)$	A or B the more stringent
(i)	Table I, p. 387	·118	·343	A
(ii)	Table III, p. 390	·070	·074	A
(iii)	Table IV, p. 391	·738	·520	B
(iv)	Table V, p. 392	·000,008	·995,351	A
(v)	Table VII, p. 396	·037	·242	A
(vi)	Table VIII, p. 397	·649	·888	A
(viii)	Table IX, p. 400	·125	·345	A
(ix)	Table XI, p. 403	·048	·976	A
(x)	Table XII, p. 405	·436	·393	B
(xi)	Table XIII, p. 406	·524	·431	B
(xii)	Table XIV, p. 409	·221	·653	A
Given below	Table II below	·575	·388	B

Thus in 8 out of 12 illustrations the more stringent direction of the probability integrals was selected by mere inspection. In the other 4 cases B ought to have been taken instead of A, but in none of these four cases was the difference such as to upset the judgment as to randomness deduced from A. The results, however, might modify a judgment as to one hypothesis being more or less probable than a second. The departure from randomness, represented by $Q_{\lambda_n} = \cdot 5$, may be in excess or defect and we must necessarily be as cautious of the approach of Q_{λ_n} to 1 as to 0. Hence in cases of doubt it may be desirable to work out the values of Q_{λ_n} by B as well as A.

A further illustration of the method may be made here, namely to testing whether a series of *Regression Coefficients* may be taken as random samples from a bivariate normal population. We reproduce the theoretical discussion and its application to a particular example which has already been given in the *Annals of Eugenics*, as it is likely to escape many readers of *Biometrika*.

The distribution of the regression coefficient $R_{y:x}$, which we will write R_t for the tth Age Group, is in samples of n_t given by*

$$z = z_0 a^{n_t - 1} \frac{1}{\{a^2 + (R_t - b)^2\}^{\frac{1}{2} n_t}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(i)},$$

where a and b are constants of the parent population, of which we are in ignorance.

* *R.S. Proc.* Vol. 112 A, pp. 1—14, or *Tables for Statisticians*, Part II, p. cli.

If we suppose a and b to be the same for all the sampled populations, then we have the combined surface in the case of k samples, if $S(n_t) = N$,

$$z = z_0' a^{N-k} \prod_{t=1}^{t=k} \frac{1}{\{a^2 + (R_t - b)^2\}^{\frac{1}{2}n_t}} \quad \text{.....(ii).}$$

To determine the most probable values of the constants a and b we must make z a maximum, or taking logarithmic differentials of z with regard to a and b:

$$\left.\begin{aligned} \frac{N-k}{a} - \sum_{t=1}^{t=k} \frac{n_t a}{a^2 + (R_t - b)^2} &= 0 \\ \sum_{t=1}^{t=k} \frac{n_t (R_t - b)}{a^2 + (R_t - b)^2} &= 0 \end{aligned}\right\} \quad \text{.....(iii).}$$

Thus to determine a^2 and b we have the equations

$$\sum_{t=1}^{t=k} \frac{n_t}{a^2 + (R_t - b)^2} = \frac{N-k}{a^2} \quad \text{.....(iv),}$$

$$\sum_{t=1}^{t=k} \frac{n_t R_t}{a^2 + (R_t - b)^2} = b \frac{N-k}{a^2} \quad \text{.....(v).}$$

These equations appear to be only capable of solution by approximation, the chief difficulty being to obtain the *first* approximation.

Returning to Equation (i) we see that if n_t be considerable, we may replace it by

$$z = \frac{z_0}{a} e^{-\frac{1}{2} n_t \left(\frac{R_t - b}{a}\right)^2} \quad \text{.....(vi),}$$

and accordingly (ii) by

$$z = \frac{z_0'}{a^k} e^{-\frac{1}{2} \sum_{t=1}^{t=k} n_t \left(\frac{R_t - b}{a}\right)^2} \quad \text{.....(vii).}$$

Again taking logarithmic differentials we have

$$b = \sum_{t=1}^{t=k} \frac{(n_t R_t)}{N} \quad \text{.....(viii),}$$

$$a^2 = \frac{N}{k} \sum_{t=1}^{t=k} \frac{n_t (R_t - b)^2}{N} \quad \text{.....(ix).}$$

Thus the first approximation to b is given by $b_1 =$ the weighted mean of R_t's and the first approximation to a^2 is a_1^2 equal N/k times the weighted standard deviation of the R_t's. If we now put $b = b_1 + \epsilon_1$ and $a^2 = a_1^2 + \epsilon_1'$ in (iv) and (v), where ϵ_1 and ϵ_1' are small quantities, the squares and products of which may be neglected, we can proceed to a second approximation and if needful to higher approximations. Substituting these values of a and b in (i) we obtain by aid of the *Tables of the Incomplete* B-*Function* the probability integral p_t of each R_t.

Illustration. Regression Coefficients of Sitting Height on Stature. As the method of determining whether a number of regression coefficients may be considered as due to random sampling from one and the same population is of interest from the

TABLE II. *To test Regression Coefficients of Sitting Height on Stature.*

(i) Age Groups	(ii) n_t	(iii) R_t	(iv) $R_t - \overline{R}_t$	(v) $(R_t - \overline{R}_t)^2$	(vi) $a/\sqrt{n_t}$	(vii) $\frac{R_t - \overline{R}_t}{a/\sqrt{n_t}}$	(viii) p_t	(ix) $\log_{10} p_t$
6–12	99	·6261	− ·2101	·0441,4201	·0805,8345⁻	− 2·61	·99534	$\bar{1}$·997,9715
13–15	331	·8554	+ ·0192	·0003,6864	·0440,7067	+ 0·44	·32997	$\bar{1}$·518,4745⁻
16	238	·8515	+ ·0153	·0002,3409	·0519,7268	+ 0·29	·38591	$\bar{1}$·586,4860
17	288	·8120	− ·0242	·0005,8564	·0472,4625⁺	− 0·51	·69497	$\bar{1}$·841,9661
18	319	·9019	+ ·0657	·0043,1649	·0448,9194	+ 1·46	·07214⁵	$\bar{2}$·858,2062
19	310	·7371	− ·0991	·0098,2081	·0455,3891	− 2·18	·98537	$\bar{1}$·993,5993
20	388	·8744	+ ·0382	·0014,5924	·0407,0497	+ 0·94	·17361	$\bar{1}$·239,5747
21	321	·9271	+ ·0909	·0082,6281	·0447,5187	+ 2·03	·02118	$\bar{2}$·345,9260
22	289	·7834	− ·0528	·0027,8784	·0471,6439	− 1·12	·86864	$\bar{1}$·938,8398
23	315	·8515	+ ·0153	·0002,3409	·0451,7602	+ 0·34	·36693	$\bar{1}$·564,5832
24	275	·7743	− ·0619	·0038,3161	·0483,5009	− 1·28	·89973	$\bar{1}$·954,1122
25	224	·8607	+ ·0245	·0006,0025	·0535,7218	+ 0·46	·32276	$\bar{1}$·508,8797
26	224	·8068	− ·0294	·0008,6436	·0535,7218	− 0·55	·70884	$\bar{1}$·850,5482
27	193	·8915	+ ·0553	·0030,5809	·0577,1449	+ 0·96	·16853	$\bar{1}$·226,6772
28	188	·8242	− ·0120	·0001,4400	·0584,7692	− 0·20⁵⁺	·58121	$\bar{1}$·764,3331
29	186	·8704	+ ·0342	·0011,6964	·0587,9050⁻	+ 0·58	·28096	$\bar{1}$·448,6445
30	197	·8445	+ ·0083	·0000,6889	·0571,2549	+ 0·14⁵⁺	·44236	$\bar{1}$·645,7758
31–32	305	·8328	− ·0034	·0000,1156	·0459,1062	− 0·07	·52790	$\bar{1}$·722,5517
33–34	280	·8441	+ ·0079	·0000,6241	·0479,1639	+ 0·16⁵⁺	·43447	$\bar{1}$·637,9598
35–36	269	·7660	− ·0702	·0049,2804	·0488,8633	− 1·44	·92507	$\bar{1}$·966,1746
37–38	265	·8288	− ·0074	·0000,5476	·0492,5387	− 0·15⁺	·55962	$\bar{1}$·747,8932
39–41	318	·8361	− ·0001	·0000,0001	·0449,6242	− 0·00	·50000	$\bar{1}$·698,9700
42–44	267	·8255	− ·0107	·0001,1449	·0490,6906	− 0·22	·58706	$\bar{1}$·768,6825
45–47	196	·8699	+ ·0337	·0011,3569	·0572,7110	+ 0·59	·27760	$\bar{1}$·443,4195
48–51	221	·8059	− ·0303	·0009,1809	·0539,3459	− 0·56	·71226	$\bar{1}$·852,6386
52–55	146	·8732	+ ·0370	·0013,6900	·0663,5704	+ 0·56	·28774	$\bar{1}$·459,0002
56–61	186	·8819	+ ·0457	·0020,8849	·0587,9050⁻	+ 0·78	·21770	$\bar{1}$·337,8584
62–81	161	·8645	+ ·0283	·0008,0089	·0631,9027	+ 0·45⁻	·32636	$\bar{1}$·513,6969
	6999	5852·6910		$S n_t (R_t - \overline{R}_t)^2 = 18{\cdot}0005{,}0868$			$\log_{10}\lambda_n = -11{\cdot}5865{,}5666$	
	$\frac{S(n_t R_t)}{N} = \overline{R}_t = {\cdot}8362$			$a^2 = {\cdot}6428{,}7531$			$\sqrt{n}\log_{10} e = 2{\cdot}298{,}0705$	
				$a = {\cdot}8017{,}9505{,}5^-$			$I\,(27,\ 5{\cdot}041{,}8630) = {\cdot}42451$	

practical standpoint, we may illustrate it on the data for the 28 Age Groups in the regression of Sitting Height on Stature. The necessary working is given at length in Table II. The numbers are here sufficiently large to justify for present purposes the use of a normal curve (vii) for the distribution of $R_{Sh:St}$. Column (i) indicates the Age Groups, (ii) the number, n_t, of cases in the sample, (iii) the observed value of $R_{Sh:St}$, (iv) the difference of this from the weighted mean value $\bar{R}_{Sh:St}$, (v) the squares of the entries in (iv), to ascertain the weighted standard deviation of $R_{Sh:St}$, (vi) gives this weighted standard deviation a divided by $\sqrt{n_t}$*, (vii) gives the ratio of the deviation in (iv) to the standard deviation $a/\sqrt{n_t}$ of $R_{Sh:St}$ in (vi), (viii) gives the values of p_t and finally (ix) the values of $\log_{10} p_t$. The sum of the values in this column (ix) is $-11{\cdot}5865{,}5666$ and since $\sqrt{n}\log_{10} e = \sqrt{28}\log_{10} e = 2{\cdot}298{,}0705$, $u = 5{\cdot}041{,}8630$ and $I(27, u) = {\cdot}42451$ from the *Incomplete* Γ-*Function Table*, using third differences. Thus the probability on our hypothesis of more divergent sets of samples of 28 Age Groups occurring is ·57549 or some 58% of sets would exhibit greater divergence. As far as this test is concerned it seems not unreasonable to suppose that the regression coefficient of Sitting Height on Stature is constant for all ages.

We may now examine this problem supposing the probability integrals measured in the reverse direction.

The values of $1 - p_t$ and their logarithms are given in Table III below:

TABLE III.

Probability Integrals of the Regression Coefficients of Sitting Height on Stature.

No.	Value of $1-p_t$	$\log_{10}(1-p_t)$	No.	Value of $1-p_t$	$\log_{10}(1-p_t)$
1	·00466	$\bar{3}{\cdot}668{,}3859$	15	·41879	$\bar{1}{\cdot}621{,}9963$
2	·67003	$\bar{1}{\cdot}826{,}0942$	16	·71904	$\bar{1}{\cdot}856{,}7531$
3	·61409	$\bar{1}{\cdot}788{,}2320$	17	·55764	$\bar{1}{\cdot}746{,}3539$
4	·30503	$\bar{1}{\cdot}484{,}3426$	18	·47210	$\bar{1}{\cdot}674{,}0340$
5	·92785	$\bar{1}{\cdot}967{,}4801$	19	·56553	$\bar{1}{\cdot}752{,}4556$
6	·01463	$\bar{2}{\cdot}165{,}2443$	20	·07493	$\bar{2}{\cdot}874{,}6557$
7	·82639	$\bar{1}{\cdot}917{,}1851$	21	·44038	$\bar{1}{\cdot}643{,}8276$
8	·97882	$\bar{1}{\cdot}990{,}7028$	22	·50000	$\bar{1}{\cdot}698{,}9700$
9	·13136	$\bar{1}{\cdot}118{,}4631$	23	·41294	$\bar{1}{\cdot}615{,}8870$
10	·63307	$\bar{1}{\cdot}801{,}4517$	24	·72240	$\bar{1}{\cdot}858{,}7777$
11	·10027	$\bar{1}{\cdot}001{,}1710$	25	·28774	$\bar{1}{\cdot}459{,}0002$
12	·67724	$\bar{1}{\cdot}830{,}7426$	26	·71226	$\bar{1}{\cdot}852{,}6386$
13	·29116	$\bar{1}{\cdot}464{,}1317$	27	·78230	$\bar{1}{\cdot}893{,}3733$
14	·83147	$\bar{1}{\cdot}919{,}8466$	28	·67364	$\bar{1}{\cdot}828{,}4279$
	$\sqrt{n}\log_{10} e = 2{\cdot}298{,}0705$			$-\log_{10}\lambda_n = 12{\cdot}679{,}3754$ $u = 5{\cdot}041{,}4005$	

* $\frac{1}{\sqrt{n_t}}$ is found at once from the tabled values in Comrie's edition of *Barlow's Tables*.

Hence $$P_{\lambda_n} = I(27, 5{\cdot}517{,}4005)$$
$$= {\cdot}6124,$$

from the *Incomplete* Γ-*Function Table.* And accordingly we may take $Q_{\lambda_n} = {\cdot}388$, which is somewhat more stringent than the $Q_{\lambda_n} = {\cdot}575$ of the A method as noted in Table I, but would not alter our judgment that the hypothesis that all the regression coefficients may be taken as due to random sampling from the same normal population of regression coefficient $R_{Sh:St} = {\cdot}8362$ is not unreasonable. Had we asked whether Case No. 1 of regression coefficient $R_{Sh:St} = {\cdot}6261$ was a reasonable sample from a population of $R_{Sh:St} = {\cdot}8362$ we should have answered in the negative. This indicates how different our judgment may be when we are considering one or twenty-eight samples. Table IV below gives the value of P_{λ_n} for given values of n, the number of probability integrals, and of $-\log_{10}\lambda_n$, i.e. the logarithm of the product of the probability integrals to base 10 with the minus sign before it. This table saves the trouble of computing $u = \dfrac{-\log_{10}\lambda_n}{\sqrt{n}\log_{10}e}$ and then interpolating into the *Incomplete* Γ-*Function Table.* The latter must still be used for values $n > 30$.

The table will suffice by using $\delta^2 P_{\lambda_n}$ to obtain values of P_{λ_n} to four figures except when n and $-\log_{10}\lambda_n$ are very small, say $n < 6$ and $-\log_{10}\lambda_n < 1{\cdot}125$, when $\delta^4 P_{\lambda_n}$ may be found from $\delta^2 P_{\lambda_n}$ if we desire four-figure accuracy.

The present Table IV may be used to find without much extra labour $P(\chi^2, n')$. It is true that it will only give the value of $P(\chi^2, n')$ to four figures, but this is usually ample, and it doubles the range of the table in *Tables for Statisticians*, going up to $n' = 61$. All we need to do is to enter the present table with

$$n = \tfrac{1}{2}(n' - 1), \text{ and } -\log_{10}\lambda_n = \tfrac{1}{2}\chi^2 \log_{10} e,$$

and then $$P(\chi^2, n') = 1 - P_{\lambda_n}.$$

For example: What is the value of $P(\chi^2, n')$ for $\chi^2 = 14$, $n' = 17$?

Here $n = 8$, $-\log_{10}\lambda_n = 7 \times \log_{10} e = 7 \times {\cdot}434{,}2944 = 3{\cdot}0406$.

But from our column for $n = 8$, we have

$$-\log_{10}\lambda_n, \text{ for } 3{\cdot}00, = {\cdot}3875,$$
$$-\log_{10}\lambda_n, \text{ for } 3{\cdot}25, = {\cdot}4729,$$

whence by linear difference the required P_{λ_n}

$$= {\cdot}3875 + {\cdot}0854 \times {\cdot}1624 = {\cdot}4014,$$

and $$P(\chi^2, n') = {\cdot}5986.$$

Palin Elderton's table gives ·5987, which is the value actually provided by the present table, if second differences are used.

Of course the *Tables of the Incomplete* Γ-*Function* carry the $P(\chi^2, n')$ Table to 7 decimal places up to $n' = 101$, but a table to 4 decimal places up to $n' = 61$ will be found not without value.

TABLE IV.

Table of the P_{λ_n} Function.

$-\log_{10}\lambda_n$	$n=2$		$n=3$		$n=4$		$n=5$		$n=6$		$-\log_{10}\lambda_n$	$n=4$ (*cont.*)		$n=5$ (*cont.*)		$n=6$ (*cont.*)	
	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$		P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$
·000	·0000	+818	·0000	+ 7	·0000	−12	·0000	—	·0000	—	7·500	·9999^{7}	—	·9998^{5}	—	·9994	—
·125	·0343	+454	·0032	+144	·0002	+24	·0000	+ 3	·0000	—	7·625	·9999^{7}	—	·9999^{-}	—	·9995^{5}	—
·250	·1139	+206	·0208	+186	·0029	+62	·0003	+13	·0000	+ 2	7·750	·9999^{8}	—	·9999^{0}	—	·9996	—
·375	·2142	+53	·0570	+170	·0117	+91	·0020	+30	·0003	+ 7	7·875	·9999^{8}	—	·9999^{2}	—	·9997	—
·500	·3127	−36	·1101	+128	·0297	+105	·0065^{+}	+47	·0012	+16	8·000	·9999^{9}	—	·9999^{4}	—	·9998	—
·675	·4216	−84	·1760	+80	·0582	+105	·0158	+62	·0037	+25	8·125	·9999^{9}	—	·9999^{5}	—	·9998	—
·750	·5151	−106	·2499	+36	·0973	+92	·0314	+71	·0086	+36	8·250	·9999^{9}	—	·9999^{6}	—	·9998^{5}	—
·875	·5980	−111	·3273	− 9	·1456	+73	·0540	+74	·0171	+45	8·375	·9999^{9}	—	·9999^{7}	—	·9999^{-}	—
1·000	·6697	−107	·4046	−28	·2012	+52	·0841	+71	·0301	+52	8·500	1·0000^{-}	—	·9999^{8}	—	·9999^{0}	—
1·125	·7308	−99	·4792	−47	·2619	+28	·1212	+63	·0483	+56	8·625			·9999^{8}	—	·9999^{2}	—
1·250	·7819	−87	·5490	−58	·3255^{+}	+ 8	·1647	+52	·0721	+56	8·750			·9999^{8}	—	·9999^{4}	—
1·375	·8243	−76	·6130	−64	·3899	− 9	·2134	+38	·1016	+53	8·875			·9999^{9}	—	·9999^{5}	—
1·500	·8592	−64	·6705^{+}	−66	·4534	−23	·2659	+24	·1364	+48	9·000			·9999^{9}	—	·9999^{6}	—
1·625	·8876	−54	·7216	−64	·5145^{+}	−34	·3208	+11	·1759	+40	9·125			·9999^{9}	—	·9999^{7}	—
1·750	·9106	−45	·7662	−60	·5723	−41	·3769	− 1	·2195^{+}	+31	9·250			·9999^{9}	—	·9999^{7}	—
1·875	·9291	−37	·8048	−55	·6260	−45	·4329	−12	·2662	+22	9·375			1·0000^{-}	—	·9999^{8}	—
2·000	·9439	−30	·8379	−50	·6751	−47	·4877	−20	·3151	+12	9·500					·9999^{8}	—
2·125	·9558	−24	·8660	−44	·7196	−47	·5405^{+}	−27	·3653	+ 3	9·625					·9999^{9}	—
2·250	·9652	−19	·8898	−38	·7594	−45	·5906	−32	·4157	− 5	9·750					·9999^{9}	—
2·375	·9727	−16	·9097	−33	·7947	−43	·6376	−36	·4657	−12	9·875					·9999^{9}	—
2·500	·9786	−12	·9262	−28	·8257	−40	·6810	−36	·5145^{-}	−18	10·000					·9999^{9}	—
2·625	·9833	−10	·9400	−24	·8527	−36	·7208	−37	·5614	−23	10·125					·9999^{9}	—
2·750	·9870	− 8	·9513	−20	·8761	−33	·7569	−36	·6061	−26	10·250					1·0000^{-}	—
2·875	·9898	− 6	·9606	−17	·8961	−29	·7894	−35	·6481	−29	10·375						
3·000	·9921	− 5	·9682	−14	·9133	−26	·8184	−33	·6874	−30							
3·125	·9939	− 4	·9744	−12	·9279	−23	·8441	−31	·7236	−30							
3·250	·9952	− 3	·9795^{-}	−10	·9402	−20	·8667	−28	·7567	−30							
3·375	·9963	− 2	·9836	− 8	·9506	−17	·8865^{+}	−26	·7869	−29							
3·500	·9971	− 2	·9869	− 6	·9593	−14	·9037	−23	·8141	−28							
3·625	·9978	− 1	·9895^{+}	− 5	·9665^{+}	−12	·9186	−21	·8385^{+}	−27							
3·750	·9983	− 1	·9917	− 4	·9726	− 9	·9314	−18	·8603	−25							
3·875	·9987	− 1	·9934	− 3	·9776	− 9	·9424	−16	·8795^{+}	−23							
4·000	·9990	− 1	·9947	− 3	·9817	− 7	·9517	−14	·8965^{-}	−21							
4·125	·9992	− 1	·9958	− 2	·9851	− 6	·9597	−12	·9114	−19							
4·250	·9994	—	·9967	− 2	·9879	− 5	·9664	−10	·9244	−17							
4·375	·9995^{+}	—	·9974	− 1	·9902	− 4	·9721	− 9	·9357	−15							
4·500	·9996	—	·9979	− 1	·9921	− 3	·9769	− 8	·9454	−14							
4·625	·9997	—	·9984	− 1	·9936	− 3	·9809	− 7	·9538	−12							
4·750	·9998	—	·9987	− 1	·9948^{5}	− 2	·9842	− 6	·9605^{+}	−10							
4·875	·9998	—	·9990	− 1	·9958^{5}	− 2	·9870	− 4	·9672	− 9							
5·000	·9999	—	·9992	—	·9967	− 2	·9894	− 4	·9725^{-}	− 8							
5·125	·9999^{0}	—	·9994	—	·9973	− 1	·9913	− 3	·9770	− 7							
5·250	·9999^{3}	—	·9995^{+}	—	·9979	− 1	·9929	− 3	·9808	− 6							
5·375	·9999^{4}	—	·9996	—	·9983	− 1	·9942	− 2	·9840	− 5							
5·500	·9999^{6}	—	·9997	—	·9986	− 1	·9952	− 2	·9867	− 4							
5·625	·9999^{7}	—	·9998	—	·9989	− 1	·9961	− 2	·9889	− 4							
5·750	·9999^{7}	—	·9998	—	·9991	—	·9968	− 1	·9908	− 3							
5·875	·9999^{8}	—	·9999	—	·9993	—	·9974	− 1	·9924	− 3							
6·000	·9999^{9}	—	·9999^{0}	—	·9994^{5}	—	·9979	− 1	·9937	− 2							
6·125	1·0000^{-}	—	·9999^{1}	—	·9996	—	·9983	− 1	·9948	− 2							
6·250			·9999^{3}	—	·9996^{5}	—	·9986	− 1	·9958	− 2							
6·375			·9999^{5}	—	·9997	—	·9989	—	·9965^{+}	− 1							
6·500			·9999^{6}	—	·9998	—	·9991	—	·9971	− 1							
6·625			·9999^{7}	—	·9998	—	·9993	—	·9977	− 1							
6·750			·9999^{8}	—	·9999	—	·9994	—	·9981	− 1							
6·875			·9999^{8}	—	·9999	—	·9995^{+}	—	·9984	− 1							
7·000			·9999^{9}	—	·9999^{2}	—	·9996	—	·9987	− 1							
7·125			·9999^{9}	—	·9999^{3}	—	·9997	—	·9990	—							
7·250			1·0000^{-}	—	·9999^{5}	—	·9998	—	·9992	—							
7·375					·9999^{6}	—	·9998	—	·9993	—							

Table of the P_{λ_n} Function (cont.)

$-\log_{10}\lambda_n$	$n=7$		$n=8$		$n=9$		$n=10$		$n=11$		$n=12$		$n=13$		$n=14$		$-\log_{10}\lambda_n$
	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	
0·00	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	0·00
0·25	·0000	+ 2	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	0·25
0·50	·0002	+17	·0000	+ 4	·0000	+ 1	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	0·50
0·75	·0020	+55	·0004	+18	·0001	+ 4	·0000	+ 1	·0000	—	·0000	—	·0000	—	·0000	—	0·75
1·00	·0094	+109	·0026	+47	·0006	+18	·0001	+ 6	·0000	+ 2	·0000	+ 1	·0000	—	·0000	—	1·00
1·25	·0277	+157	·0095$^-$	+86	·0029	+39	·0008	+15	·0002	+ 5	·0001	+ 1	·0000	+ 1	·0000	—	1·25
1·50	·0618	+179	·0250$^+$	+124	·0091	+69	·0030	+33	·0009	+14	·0003	+ 5	·0001	+ 1	·0000	—	1·50
1·75	·1138	+169	·0529	+147	·0223	+99	·0085$^+$	+56	·0030	+27	·0010	+12	·0003	+ 5	·0000	+3	1·75
2·00	·1827	+132	·0955$^+$	+148	·0453	+121	·0196	+82	·0078	+47	·0029	+23	·0010	+10	·0003	+ 4	2·00
2·25	·2647	+ 79	·1529	+129	·0805$^+$	+129	·0389	+100	·0173	+66	·0071	+38	·0027	+20	·0010	+ 9	2·25
2·50	·3546	+ 23	·2232	+ 93	·1287	+121	·0682	+112	·0334	+84	·0151	+55	·0064	+31	·0025$^+$	+16	2·50
2·75	·4469	− 27	·3029	+ 50	·1889	+ 99	·1087	+111	·0579	+ 97	·0287	+71	·0132	+46	·0057	+26	2·75
3·00	·5365$^-$	− 65	·3875$^+$	+ 7	·2590	+ 67	·1603	+ 98	·0921	+100	·0493	+84	·0246	+60	·0116	+38	3·00
3·25	·6196	− 89	·4729	− 31	·3358	+ 32	·2217	+ 76	·1363	+ 95	·0783	+89	·0421	+72	·0212	+51	3·25
3·50	·6938	−101	·5552	− 59	·4157	− 3	·2907	+ 49	·1900	+ 80	·1162	+90	·0667	+80	·0360	+62	3·50
3·75	·7579	−103	·6316	− 78	·4953	− 33	·3646	+ 18	·2517	+ 58	·1631	+80	·0993	+82	·0569	+71	3·75
4·00	·8117	− 98	·7002	− 88	·5717	− 55	·4403	− 10	·3192	+ 35	·2179	+65	·1401	+78	·0849	+75	4·00
4·25	·8558	− 88	·7599	− 90	·6426	− 70	·5150$^+$	− 33	·3902	+ 9	·2792	+45	·1886	+68	·1204	+74	4·25
4·50	·8911	− 76	·8106	− 86	·7064	− 78	·5864	− 52	·4621	− 14	·3450$^+$	+24	·2439	+53	·1633	+68	4·50
4·75	·9188	− 63	·8527	− 79	·7624	− 80	·6526	− 63	·5326	− 34	·4132	+ 2	·3045$^-$	+35	·2129	+57	4·75
5·00	·9401	− 51	·8870	− 69	·8104	− 77	·7125$^-$	− 71	·5997	− 48	·4817	− 17	·3685$^+$	+16	·2683	+43	5·00
5·25	·9564	− 41	·9143	− 59	·8507	− 71	·7653	− 73	·6620	− 59	·5485$^-$	− 34	·4342	− 2	·3278	+27	5·25
5·50	·9685$^+$	− 32	·9357	− 49	·8839	− 63	·8108	− 68	·7184	− 65	·6119	− 46	·4995$^+$	− 19	·3901	+10	5·50
5·75	·9775$^+$	− 24	·9524	− 39	·9107	− 55	·8495$^-$	− 66	·7683	− 65	·6707	− 54	·5630	− 33	·4533	− 6	5·75
6·00	·9841	− 18	·9650$^+$	− 31	·9321	− 46	·8816	− 58	·8117	− 63	·7241	− 59	·6231	− 44	·5159	− 21	6·00
6·25	·9888	− 13	·9746	− 24	·9489	− 38	·9079	− 51	·8488	− 60	·7716	− 60	·6789	− 51	·5764	− 33	6·25
6·50	·9922	− 10	·9817	− 19	·9619	− 31	·9291	− 44	·8799	− 54	·8130	− 59	·7296	− 55	·6336	− 42	6·50
6·75	·9946	− 8	·9869	− 14	·9719	− 24	·9459	− 35	·9056	− 48	·8486	− 55	·7749	− 56	·6867	− 48	6·75
7·00	·9962	− 5	·9907	− 11	·9794	− 19	·9592	− 31	·9265$^+$	− 41	·8788	− 51	·8146	− 54	·7350$^-$	− 51	7·00
7·25	·9975$^-$	− 3	·9934	− 8	·9850$^+$	− 15	·9694	− 23	·9434	− 35	·9038	− 43	·8488	− 51	·7782	− 52	7·25
7·50	·9983	− 2	·9954	− 6	·9892	− 11	·9773	− 19	·9567	− 29	·9245$^-$	− 40	·8780	− 47	·8163	− 50	7·50
7·75	·9988	− 1	·9968	− 4	·9923	− 8	·9833	− 15	·9672	− 23	·9412	− 33	·9025$^-$	− 42	·8493	− 48	7·75
8·00	·9992	− 1	·9978	− 3	·9945$^+$	− 6	·9878	− 12	·9754	− 19	·9546	− 28	·9228	− 37	·8776	− 44	8·00
8·25	·9995$^-$	—	·9985$^-$	− 2	·9961	− 5	·9911	− 8	·9816	− 15	·9653	− 23	·9393	− 32	·9015$^-$	− 40	8·25
8·50	·9997	—	·9990	− 2	·9973	− 4	·9936	− 7	·9864	− 12	·9736	− 19	·9528	− 27	·9214	− 35	8·50
8·75	·9998	—	·9993	− 1	·9981	− 3	·9954	− 5	·9900	− 9	·9801	− 15	·9635$^+$	− 22	·9378	− 30	8·75
9·00	·9998	—	·9995$^+$	—	·9987	− 2	·9967	− 4	·9927	− 7	·9851	− 12	·9720	− 18	·9512	− 26	9·00
9·25	·9999$^-$	—	·9997	—	·9991	− 1	·9977	− 3	·9947	− 5	·9889	− 9	·9787	− 15	·9620	− 22	9·25
9·50	·9999^3	—	·9998	—	·9994	− 1	·9984	− 2	·9962	− 4	·9918	− 7	·9839	− 12	·9706	− 18	9·50
9·75	·9999^5	—	·9999$^-$	—	·9996	− 1	·9989	− 2	·9973	− 3	·9940	− 5	·9879	− 9	·9774	− 14	9·75
10·00	·9999^7	—	·9999^0	—	·9997	—	·9992	− 1	·9981	− 2	·9956	− 3	·9910	− 7	·9828	− 12	10·00
10·25	·9999^8	—	·9999^4	—	·9998	—	·9994	− 1	·9986	− 1	·9968	− 2	·9933	− 5	·9870	− 10	10·25
10·50	·9999^9	—	·9999^6	—	·9999$^-$	—	·9996	—	·9990	− 1	·9977	− 2	·9951	− 4	·9902	− 8	10·50
10·75	·9999^9	—	·9999^7	—	·9999^1	—	·9997	—	·9993	− 1	·9983	− 1	·9964	− 3	·9927	− 7	10·75
11·00	1·0000$^-$	—	·9999^8	—	·9999^4	—	·9998	—	·9995$^+$	—	·9988	− 1	·9973	− 2	·9945$^+$	− 5	11·00
11·25			·9999^9	—	·9999^6	—	·9999$^-$	—	·9997	—	·9992	− 1	·9981	− 1	·9960	− 4	11·25
11·50			·9999^9	—	·9999^7	—	·9999^2	—	·9998	—	·9994	—	·9985$^+$	− 1	·9970	− 3	11·50
11·75			1·0000$^-$	—	·9999^8	—	·9999^4	—	·9998	—	·9996	—	·9990	− 1	·9978	− 2	11·75
12·00					·9999^9	—	·9999^6	—	·9999$^-$	—	·9997	—	·9993	—	·9984	− 1	12·00
12·25					·9999^9	—	·9999^7	—	·9999^3	—	·9998	—	·9995$^-$	—	·9989	− 1	12·25
12·50					·9999^9	—	·9999^8	—	·9999^5	—	·9999$^-$	—	·9996	—	·9992	− 1	12·50
12·75					1·0000$^-$	—	·9999^9	—	·9999^7	—	·9999^0	—	·9997	—	·9994	− 1	12·75
13·00							·9999^9	—	·9999^8	—	·9999^3	—	·9998	—	·9996	—	13·00
13·25							1·0000$^-$	—	·9999^8	—	·9999^5	—	·9999$^-$	—	·9997	—	13·25
13·50									·9999^9	—	·9999^7	—	·9999^1	—	·9998	—	13·50
13·75									·9999^9	—	·9999^8	—	·9999^4	—	·9998	—	13·75
14·00									1·0000$^-$	—	·9999^9	—	·9999^6	—	·9999$^-$	—	14·00
14·25											·9999^9	—	·9999^7	—	·9999^3	—	14·25
14·50											·9999^9	—	·9999^8	—	·9999^5	—	14·50
14·75											1·0000$^-$	—	·9999^9	—	·9999^6	—	14·75
15·00													·9999^9	—	·9999^7	—	15·00
15·25													·9999^9	—	·9999^8	—	15·25
15·50													1·0000$^-$	—	·9999^9	—	15·50
15·75															·9999^9	—	15·75
16·00															·9999^9	—	16·00
16·25															1·0000$^-$	—	16·25

Table of the P_{λ_n} Function (cont.)

$-\log_{10}\lambda_n$	$n=15$		$n=16$		$n=17$		$n=18$		$n=19$		$n=20$		$n=21$		$n=22$		$-\log_{10}\lambda_n$
	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	
0·00																	
0·25																	
0·50																	
0·75																	
1·00																	
1·25																	
1·50	1·50	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	1·50
1·75	1·75	·0000	+ 1	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	1·75
2·00	2·00	·0000	+ 1	·0000	+ 1	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	2·00
2·25	2·25	·0003	+ 4	·0001	+ 1	·0000	+ 1	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	2·25
2·50	2·50	·0009	+ 8	·0003	+ 4	·0001	+ 1	·0000	+ 1	·0000	—	·0000	—	·0000	—	·0000	—	2·50
2·75	2·75	·0023	+14	·0009	+ 7	·0003	+ 3	·0001	+ 1	·0000	+ 1	·0000	—	·0000	—	·0000	—	2·75
3·00	3·00	·0051	+22	·0021	+12	·0008	+ 6	·0003	+ 3	·0001	+ 1	·0000	+ 1	·0000	—	·0000	—	3·00
3·25	3·25	·0101	+32	·0045$^+$	+19	·0019	+10	·0008	+ 4	·0003	+ 2	·0001	+ 1	·0000	+1	·0000	—	3·25
3·50	3·50	·0183	+43	·0088	+27	·0040	+16	·0017	+ 8	·0007	+ 4	·0003	+ 2	·0001	+ 1	·0000	+ 1	3·50
3·75	3·75	·0308	+54	·0158	+37	·0076	+23	·0035$^+$	+13	·0015$^+$	+ 8	·0007	+ 3	·0003	+ 1	·0001	—	3·75
4·00	4·00	·0487	+62	·0264	+46	·0136	+31	·0067	+20	·0031	+11	·0014	+ 6	·0006	+ 3	·0002	+ 2	4·00
4·25	4·25	·0728	+68	·0417	+55	·0227	+40	·0117	+27	·0058	+16	·0027	+10	·0012	+ 6	·0005$^+$	+ 3	4·25
4·50	4·50	·1037	+71	·0625$^-$	+61	·0358	+48	·0195$^+$	+35	·0101	+24	·0050$^+$	+14	·0024	+ 8	·0011	+ 4	4·50
4·75	4·75	·1414	+69	·0893	+64	·0537	+54	·0307	+42	·0168	+29	·0088	+20	·0044	+12	·0021	+ 7	4·75
5·00	5·00	·1858	+66	·1225$^+$	+63	·0770	+59	·0461	+49	·0264	+37	·0145$^-$	+26	·0076	+17	·0038	+10	5·00
5·25	5·25	·2360	+59	·1621	+59	·1062	+60	·0664	+53	·0397	+44	·0227	+32	·0125$^-$	+22	·0066	+14	5·25
5·50	5·50	·2911	+48	·2075$^-$	+51	·1413	+57	·0920	+56	·0574	+47	·0342	+38	·0196	+28	·0108	+19	5·50
5·75	5·75	·3496	+34	·2580	+40	·1822	+52	·1232	+55	·0798	+52	·0496	+44	·0295$^+$	+34	·0169	+25	5·75
6·00	6·00	·4100	+20	·3125$^+$	+28	·2283	+44	·1599	+52	·1074	+52	·0692	+47	·0428	+39	·0255$^-$	+30	6·00
6·25	6·25	·4709	+ 4	·3698	+14	·2788	+34	·2018	+46	·1402	+52	·0936	+49	·0601	+43	·0371	+35	6·25
6·50	6·50	·5309	−21	·4285$^-$	+ 1	·3327	+22	·2483	+38	·1782	+46	·1229	+49	·0816	+46	·0521	+39	6·50
6·75	6·75	·5888	−32	·4873	−12	·3887	+10	·2986	+28	·2208	+41	·1572	+47	·1077	+47	·0711	+43	6·75
7·00	7·00	·6434	−40	·5449	−23	·4458	− 2	·3517	+17	·2675$^+$	+33	·1961	+42	·1385$^+$	+46	·0944	+44	7·00
7·25	7·25	·6940	−45	·6003	−31	·5025$^+$	−13	·4065$^+$	+ 6	·3175$^-$	+22	·2392	+36	·1739	+43	·1220	+44	7·25
7·50	7·50	·7402	−48	·6526	−38	·5580	−23	·4619	− 5	·3697	+14	·2859	+28	·2136	+38	·1541	+43	7·50
7·75	7·75	·7816	−48	·7010	−42	·6111	−31	·5168	−15	·4233	+ 2	·3354	+19	·2570	+31	·1904	+39	7·75
8·00	8·00	·8182	−47	·7452	−45	·6612	−36	·5702	−23	·4771	− 7	·3868	+ 9	·3036	+23	·2306	+34	8·00
8·25	8·25	·8501	−45	·7850$^-$	−45	·7077	−40	·6213	−30	·5302	−16	·4391	—	·3525$^+$	+15	·2742	+27	8·25
8·50	8·50	·8775$^-$	−41	·8202	−44	·7501	−42	·6694	−35	·5817	−23	·4914	− 9	·4029	+ 6	·3205$^+$	+20	8·50
8·75	8·75	·9007	−37	·8510	−42	·7883	−43	·7141	−38	·6309	−29	·5428	−17	·4540	− 2	·3688	+12	8·75
9·00	9·00	·9203	−33	·8776	−39	·8223	−42	·7549	−40	·6772	−33	·5926	−24	·5049	−10	·4183	+ 4	9·00
9·25	9·25	·9364	−29	·9002	−35	·8520	−40	·7916	−40	·7202	−37	·6400	−27	·5547	−17	·4682	− 4	9·25
9·50	9·50	·9498	−25	·9193	−31	·8778	−37	·8244	−39	·7594	−38	·6847	−33	·6029	−23	·5177	−11	9·50
9·75	9·75	·9606	−21	·9353	−28	·9000	−34	·8532	−38	·7949	−38	·7260	−35	·6487	−28	·5661	−17	9·75
10·00	10·00	·9693	−18	·9485$^+$	−24	·9187	−30	·8783	−35	·8266	−37	·7639	−36	·6918	−31	·6127	−23	10·00
10·25	10·25	·9762	−15	·9593	−20	·9344	−27	·8998	−32	·8545$^-$	−35	·7982	−36	·7317	−33	·6570	−27	10·25
10·50	10·50	·9817	−12	·9681	−17	·9475$^-$	−23	·9182	−29	·8789	−34	·8288	−36	·7682	−35	·6986	−30	10·50
10·75	10·75	·9860	−10	·9751	−14	·9583	−20	·9337	−26	·8999	−30	·8558	−34	·8014	−35	·7371	−32	10·75
11·00	11·00	·9894	− 8	·9808	−12	·9670	−17	·9466	−22	·9179	−28	·8795$^+$	−32	·8310	−34	·7725$^-$	−33	11·00
11·25	11·25	·9920	− 6	·9852	−10	·9741	−14	·9573	−19	·9331	−24	·9001	−29	·8573	−32	·8045$^+$	−33	11·25
11·50	11·50	·9940	− 5	·9887	− 8	·9798	−12	·9661	−17	·9459	−22	·9177	−26	·8803	−30	·8332	−32	11·50
11·75	11·75	·9955$^+$	− 4	·9914	− 6	·9844	−10	·9732	−14	·9565$^-$	−19	·9326	−23	·9003	−28	·8587	−31	11·75
12·00	12·00	·9967	− 3	·9935$^-$	− 5	·9880	− 8	·9790	−11	·9652	−15	·9452	−21	·9176	−25	·8812	−29	12·00
12·25	12·25	·9976	− 2	·9951	− 4	·9908	− 6	·9836	− 9	·9724	−14	·9558	−18	·9323	−23	·9007	−27	12·25
12·50	12·50	·9982	− 2	·9963	− 3	·9930	− 5	·9873	− 8	·9782	−11	·9645$^-$	−15	·9447	−20	·9176	−24	12·50
12·75	12·75	·9987	− 1	·9973	− 2	·9947	− 4	·9902	− 6	·9829	−10	·9717	−13	·9551	−17	·9320	−22	12·75
13·00	13·00	·9990	− 1	·9980	− 2	·9960	− 3	·9925$^+$	− 5	·9867	− 9	·9775$^+$	−11	·9638	−15	·9443	−19	13·00
13·25	13·25	·9993	− 1	·9985$^+$	− 1	·9970	− 2	·9943	− 4	·9897	− 7	·9823	− 9	·9710	−13	·9546	−17	13·25
13·50	13·50	·9995$^+$	—	·9989	− 1	·9978	− 2	·9957	− 3	·9920	− 5	·9861	− 8	·9769	−11	·9632	−15	13·50
13·75	13·75	·9996	—	·9992	− 1	·9983	− 1	·9967	− 3	·9938	− 4	·9892	− 6	·9817	− 9	·9704	−13	13·75
14·00	14·00	·9997	—	·9994	—	·9988	− 1	·9975$^+$	− 2	·9953	− 3	·9916	− 5	·9856	− 8	·9763	−11	14·00
14·25	14·25	·9998	—	·9996	—	·9991	− 1	·9982	− 2	·9964	− 2	·9935$^+$	− 4	·9887	− 6	·9811	− 9	14·25
14·50	14·50	·9999$^-$	—	·9997	—	·9993	—	·9986	− 1	·9973	− 2	·9950$^+$	− 3	·9912	− 5	·9850$^+$	− 8	14·50
14·75	14·75	·9999^1	—	·9998	—	·9995$^-$	—	·9990	− 1	·9980	− 2	·9961	− 3	·9931	− 4	·9882	− 6	14·75
15·00	15·00	·9999^2	—	·9998	—	·9996	—	·9992	− 1	·9985$^-$	− 1	·9971	− 3	·9947	− 3	·9908	− 5	15·00
15·25	15·25	·9999^3	—	·9999$^-$	—	·9997	—	·9994	—	·9988	− 1	·9978	− 2	·9959	− 3	·9928	− 4	15·25
15·50	15·50	·9999^7	—	·9999^2	—	·9998	—	·9996	—	·9991	− 1	·9984	− 2	·9968	− 2	·9944	− 3	15·50
15·75	15·75	·9999^8	—	·9999^4	—	·9999$^-$	—	·9997	—	·9994	− 1	·9988	− 1	·9976	− 2	·9957	− 3	15·75
16·00	16·00	·9999^8	—	·9999^6	—	·9999^0	—	·9998	—	·9995$^-$	—	·9991	− 1	·9982	− 1	·9967	− 2	16·00
16·25	16·25	·9999^9	—	·9999^7	—	·9999^3	—	·9998	—	·9997	—	·9993	− 1	·9985$^+$	− 1	·9974	− 2	16·25
16·50	16·50	·9999^9	—	·9999^8	—	·9999^5	—	·9999$^-$	—	·9998	—	·9995$^-$	—	·9990	− 1	·9981	− 1	16·50
16·75	16·75	·9999^9	—	·9999^9	—	·9999^7	—	·9999^2	—	·9998	—	·9996	—	·9992	—	·9985$^+$	− 1	16·75
17·00	17·00	1·0000$^-$	—	·9999^9	—	·9999^8	—	·9999^4	—	·9999$^-$	—	·9997	—	·9994	—	·9989	− 1	17·00
17·25	17·25			·9999^9	—	·9999^9	—	·9999^6	—	·9999^0	—	·9998	—	·9996	—	·9991	− 1	17·25
16·00																		
16·25																		

Table of the P_{λ_n} *Function (cont.)*

$-\log_{10}\lambda_n$	$n=15$ (*cont.*)		$n=16$ (*cont.*)		$n=17$ (*cont.*)		$n=18$ (*cont.*)		$n=19$ (*cont.*)		$n=20$ (*cont.*)		$n=21$ (*cont.*)		$n=22$ (*cont.*)		$-\log_{10}\lambda_n$
	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	
17·25			·9999^{9}	—	·9999^{8}	—	·9999^{6}	—	·9999^{0}	—	·9998	—	·9996	—	·9991	− 1	17·25
17·50			1·0000^{-}	—	·9999^{9}	—	·9999^{7}	—	·9999^{3}	—	·9998	—	·9997	—	·9994	—	17·50
17·75					·9999^{9}	—	·9999^{8}	—	·9999^{5}	—	·9999^{-}	—	·9998	—	·9995^{+}	—	17·75
18·00					·9999^{9}	—	·9999^{8}	—	·9999^{6}	—	·9999^{2}	—	·9998	—	·9996	—	18·00
18·25					1·0000^{-}	—	·9999^{9}	—	·9999^{7}	—	·9999^{4}	—	·9999^{-}	—	·9997	—	18·25
18·50							·9999^{9}	—	·9999^{8}	—	·9999^{6}	—	·9999^{1}	—	·9998	—	18·50
18·75							·9999^{9}	—	·9999^{9}	—	·9999^{7}	—	·9999^{3}	—	·9999^{-}	—	18·75
19·00							1·0000^{-}	—	·9999^{9}	—	·9999^{8}	—	·9999^{5}	—	·9999^{-}	—	19·00
19·25									·9999^{9}	—	·9999^{8}	—	·9999^{6}	—	·9999^{2}	—	19·25
19·50									1·0000^{-}	—	·9999^{9}	—	·9999^{7}	—	·9999^{4}	—	19·50
19·75											·9999^{9}	—	·9999^{8}	—	·9999^{6}	—	19·75
20·00											·9999^{9}	—	·9999^{9}	—	·9999^{7}	—	20·00
20·25											1·0000^{-}	—	·9999^{9}	—	·9999^{8}	—	20·25
20·50													·9999^{9}	—	·9999^{8}	—	20·50
20·75													1·0000^{-}	—	·9999^{9}	—	20·75
21·00															·9999^{9}	—	21·00
21·25															·9999^{9}	—	21·25
21·50															1·0000^{-}	—	21·50

$-\log_{10}\lambda_n$	$n=23$		$n=24$		$n=25$		$n=26$		$n=27$		$n=28$		$n=29$		$n=30$		$-\log_{10}\lambda_n$
	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	
3·50	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	3·50
3·75	·0000	+ 1	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	3·75
4·00	·0001	—	·0000	+ 1	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	4·00
4·25	·0002	+ 2	·0001	—	·0000	+ 1	·0000	—	·0000	—	·0000	—	·0000	—	·0000	—	4·25
4·50	·0005^{+}	+ 2	·0002	+ 1	·0001	—	·0000	+ 1	·0000	—	·0000	—	·0000	—	·0000	—	4·50
4·75	·0010	+ 3	·0004	+ 2	·0002	+ 1	·0001	—	·0000	+ 1	·0000	—	·0000	—	·0000	—	4·75
5·00	·0018	+ 7	·0009	+ 3	·0004	+ 2	·0002	+ 1	·0001	—	·0000	+ 1	·0000	—	·0000	—	5·00
5·25	·0033	+ 9	·0016	+ 6	·0008	+ 2	·0004	—	·0002	—	·0001	—	·0000	+ 1	·0000	—	5·25
5·50	·0057	+12	·0029	+ 9	·0014	+ 5	·0006	+ 4	·0003	+ 2	·0001	+ 1	·0001	—	·0000	+ 1	5·50
5·75	·0093	+17	·0049	+11	·0025^{+}	+ 7	·0012	+ 4	·0006	+ 2	·0003	+ 2	·0001	+ 1	·0001	—	5·75
6·00	·0146	+21	·0080	+15	·0043	+10	·0022	+ 5	·0011	+ 4	·0005^{+}	+ 2	·0002	+ 2	·0001	+ 1	6·00
6·25	·0220	+26	·0126	+18	·0070	+13	·0037	+ 8	·0019	+ 5	·0010	+ 3	·0005^{-}	+ 2	·0002	+ 1	6·25
6·50	·0321	+31	·0190	+24	·0109	+16	·0060	+11	·0032	+ 7	·0017	+ 4	·0008	+ 3	·0004	+ 2	6·50
6·75	·0453	+36	·0278	+27	·0165^{-}	+20	·0094	+14	·0052	+ 9	·0028	+ 6	·0014	+ 4	·0008	+ 2	6·75
7·00	·0620	+39	·0393	+32	·0241	+25	·0142	+18	·0082	+12	·0045^{+}	+ 8	·0024	+ 5	·0013	+ 3	7·00
7·25	·0826	+41	·0540	+36	·0341	+29	·0209	+22	·0123	+16	·0071	+11	·0039	+ 7	·0021	+ 4	7·25
7·50	·1074	+42	·0723	+38	·0471	+33	·0297	+26	·0181	+19	·0107	+14	·0061	+ 9	·0034	+ 6	7·50
7·75	·1364	+41	·0945^{-}	+40	·0633	+36	·0411	+29	·0258	+23	·0157	+17	·0093	+12	·0053	+ 8	7·75
8·00	·1695^{+}	+39	·1206	+40	·0831	+38	·0554	+33	·0358	+27	·0224	+21	·0136	+15	·0080	+11	8·00
8·25	·2066	+35	·1508	+39	·1066	+38	·0730	+35	·0485^{-}	+30	·0312	+24	·0195^{-}	+18	·0118	+13	8·25
8·50	·2472	+30	·1848	+36	·1340	+38	·0941	+37	·0641	+33	·0424	+27	·0272	+22	·0170	+16	8·50
8·75	·2908	+24	·2225^{-}	+32	·1651	+36	·1189	+37	·0831	+34	·0563	+30	·0371	+25	·0237	+19	8·75
9·00	·3368	+17	·2633	+27	·1999	+33	·1473	+36	·1054	+35	·0733	+32	·0495^{-}	+28	·0324	+22	9·00
9·25	·3844	+ 9	·3069	+21	·2380	+29	·1794	+34	·1313	+35	·0934	+34	·0646	+30	·0434	+25	9·25
9·50	·4340	+ 1	·3524	+14	·2790	+24	·2148	+30	·1607	+34	·1169	+34	·0827	+32	·0569	+28	9·50
9·75	·4817	− 5	·3994	+ 7	·3224	+18	·2532	+26	·1935^{-}	+31	·1438	+33	·1040	+33	·0732	+30	9·75
10·00	·5299	−12	·4470	—	·3675^{-}	+11	·2942	+21	·2294	+28	·1741	+32	·1286	+32	·0925^{-}	+31	10·00
10·25	·5768	−18	·4946	− 7	·4137	+ 5	·3373	+15	·2680	+23	·2074	+29	·1564	+31	·1148	+31	10·25
10·50	·6220	−23	·5415^{-}	−13	·4604	− 2	·3819	+ 9	·3090	+18	·2437	+25	·1873	+29	·1403	+31	10·50
10·75	·6648	−27	·5871	−18	·5069	− 8	·4274	+ 3	·3518	+12	·2825^{-}	+21	·2212	+27	·1689	+30	10·75
11·00	·7051	−29	·6309	−22	·5525^{+}	−14	·4732	− 3	·3959	+ 7	·3234	+16	·2578	+23	·2005^{-}	+27	11·00
11·25	·7424	−31	·6724	−26	·5969	−18	·5186	− 9	·4406	+ 1	·3658	+10	·2966	+19	·2348	+24	11·25

Table of the P_{λ_n} Function (cont.)

$-\log_{10}\lambda_n$	$n=23$ (cont.)		$n=24$ (cont.)		$n=25$ (cont.)		$n=26$ (cont.)		$n=27$ (cont.)		$n=28$ (cont.)		$n=29$ (cont.)		$n=30$ (cont.)		$-\log_{10}\lambda_n$
	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	P_{λ_n}	$\delta^2 P_{\lambda_n}$	
11·25	·7424	−31	·6724	−26	·5969	−18	·5186	− 9	·4406	+ 1	·3658	+10	·2966	+19	·2348	+24	11·25
11·50	·7766	−32	·7113	−28	·6393	−22	·5631	−14	·4855⁻	− 5	·4093	+ 5	·3373	+14	·2715⁺	+21	11·50
11·75	·8076	−32	·7474	−30	·6796	−26	·6062	−18	·5298	−10	·4533	− 1	·3793	+ 9	·3103	+16	11·75
12·00	·8355⁻	−31	·7806	−30	·7173	−27	·6475⁻	−22	·5732	−14	·4972	− 6	·4223	+ 3	·3508	+12	12·00
12·25	·8603	−29	·8106	−30	·7523	−29	·6865⁺	−25	·6152	−18	·5406	−11	·4655⁻	− 2	·3924	+ 7	12·25
12·50	·8821	−28	·8377	−29	·7844	−29	·7231	−27	·6553	−22	·5829	−15	·5086	− 7	·4348	+ 2	12·50
12·75	·9012	−26	·8618	−28	·8136	−29	·7571	−28	·6932	−24	·6238	−18	·5510	−11	·4773	− 3	12·75
13·00	·9177	−23	·8831	−27	·8399	−28	·7882	−28	·7287	−26	·6628	−21	·5922	−15	·5195⁻	− 8	13·00
13·25	·9319	−21	·9017	−25	·8634	−27	·8166	−28	·7617	−27	·6996	−24	·6320	−18	·5609	−12	13·25
13·50	·9439	−19	·9179	−22⁵	·8841	−25	·8421	−27	·7919	−27	·7341	−25	·6700	−21	·6012	−15	13·50
13·75	·9541	−16	·9318	−20	·9023	−24	·8649	−26	·8195⁻	−27	·7661	−26	·7058	−23	·6400	−18	13·75
14·00	·9627	−15	·9437	−18	·9181	−22	·8852	−24	·8443	−26	·7955⁺	−26	·7393	−24	·6769	−21	14·00
14·25	·9698	−12	·9538	−16	·9318	−20	·9029	−23	·8665⁺	−25	·8223	−26	·7704	−25	·7118	−23	14·25
14·50	·9757	−10	·9623	−14	·9435⁻	−17	·9184	−21	·8863	−24	·8465⁻	−25	·7990	−25	·7444	−24	14·50
14·75	·9806	− 9	·9694	−12	·9535⁻	−15	·9319	−19	·9036	−22	·8681	−24	·8251	−25	·7746	−24	14·75
15·00	·9846	− 8	·9753	−11	·9619	−14	·9434	−17	·9188	−20	·8874	−23	·8486	−24	·8024	−24	15·00
15·25	·9878	− 7	·9801	− 8	·9689	−12	·9532	−15	·9320	−18	·9044	−21	·8697	−23	·8278	−24	15·25
15·50	·9904	− 6	·9841	− 7	·9748	−10	·9615⁺	−13	·9433	−16	·9192	−19	·8885⁺	−22	·8507	−23	15·50
15·75	·9924	− 5	·9874	− 6	·9797	− 9	·9686	−11	·9530	−15	·9322	−18	·9051	−20	·8713	−22	15·75
16·00	·9941	− 4	·9900	− 5	·9837	− 7	·9744	−10	·9613	−13	·9433	−16	·9197	−19	·8897	−21	16·00
16·25	·9954	− 3	·9921	− 4	·9870	− 6	·9793	− 8	·9682	−11	·9529	−14	·9324	−17	·9059	−20	16·25
16·50	·9964	− 2	·9938	− 3	·9897	− 5	·9833	− 7	·9741	−10	·9610	−12	·9434	−15	·9202	−18	16·50
16·75	·9972	− 1	·9952	− 3	·9918	− 4	·9866	− 6	·9789	− 8	·9679	−11	·9528	−14	·9326	−17	16·75
17·00	·9979	− 1	·9963	− 3	·9936	− 3	·9893	− 5	·9830	− 7	·9737	− 9	·9608	−12	·9434	−15	17·00
17·25	·9984	− 1	·9971	− 2	·9949	− 3	·9915	− 4	·9863	− 6	·9786	− 8	·9677	−11	·9527	−13	17·25
17·50	·9988	− 1	·9977	− 2	·9961	− 2	·9933	− 3	·9890	− 5	·9826	− 7	·9735⁻	− 9	·9607	−12	17·50
17·75	·9991	− 1	·9983	− 2	·9969	− 2	·9947	− 3	·9912	− 4	·9860	− 6	·9783	− 8	·9675⁻	−10	17·75
18·00	·9993	—	·9987	− 1	·9976	− 2	·9959	− 3	·9930	− 3	·9887	− 5	·9823	− 7	·9732	− 9	18·00
18·25	·9994	—	·9990	− 1	·9982	− 1	·9967	− 2	·9945⁺	− 3	·9910	− 4	·9857	− 6	·9780	− 8	18·25
18·50	·9996	—	·9992	—	·9986	− 1	·9974	− 2	·9957	− 2	·9928	− 3	·9885⁻	− 5	·9821	− 7	18·50
18·75	·9997	—	·9994	—	·9989	—	·9980	− 1	·9966	− 2	·9943	− 3	·9907	− 4	·9854	− 6	18·75
19·00	·9998	—	·9996	—	·9992	—	·9985⁻	− 1	·9973	− 1	·9955⁻	− 3	·9926	− 3	·9882	− 5	19·00
19·25	·9998	—	·9997	—	·9994	—	·9988	− 1	·9979	− 1	·9964	− 2	·9941	− 3	·9905⁻	− 4	19·25
19·50	·9999⁻	—	·9998	—	·9995⁺	—	·9991	− 1	·9984	− 1	·9972	− 2	·9953	− 2	·9924	− 3	19·50
19·75	·9999⁰	—	·9998	—	·9996	—	·9994	—	·9987	− 1	·9978	− 1	·9963	− 2	·9939	− 3	19·75
20·00	·9999⁴	—	·9999⁻	—	·9997	—	·9995⁻	—	·9990	—	·9983	− 1	·9971	− 2	·9951	− 2	20·00
20·25	·9999⁵	—	·9999⁰	—	·9998	—	·9996	—	·9993	—	·9987	− 1	·9977	− 1	·9961	− 2	20·25
20·50	·9999⁷	—	·9999³	—	·9999⁻	—	·9997	—	·9994	—	·9990	− 1	·9982	− 1	·9970	− 2	20·50
20·75	·9999⁸	—	·9999⁵	—	·9999⁻	—	·9998	—	·9996	—	·9992	—	·9986	− 1	·9976	− 1	20·75
21·00	·9999⁸	—	·9999⁶	—	·9999²	—	·9998	—	·9997	—	·9994	—	·9989	− 1	·9981	− 1	21·00
21·25	·9999⁹	—	·9999⁷	—	·9999⁴	—	·9999⁻	—	·9998	—	·9995⁺	—	·9992	—	·9985⁺	− 1	21·25
21·50	·9999⁹	—	·9999⁸	—	·9999⁶	—	·9999¹	—	·9998	—	·9996	—	·9994	—	·9989	− 1	21·50
21·75	·9999⁹	—	·9999⁸	—	·9999⁷	—	·9999³	—	·9999⁻	—	·9997	—	·9995⁻	—	·9991	—	21·75
22·00	·9999⁹	—	·9999⁹	—	·9999⁸	—	·9999⁵	—	·9999⁰	—	·9998	—	·9996	—	·9993	—	22·00
22·25	1·0000⁻	—	·9999⁹	—	·9999⁸	—	·9999⁶	—	·9999²	—	·9998	—	·9997	—	·9995⁻	—	22·25
22·50			·9999⁹	—	·9999⁹	—	·9999⁷	—	·9999⁴	—	·9999⁻	—	·9998	—	·9996	—	22·50
22·75			1·0000⁻	—	·9999⁹	—	·9999⁸	—	·9999⁶	—	·9999¹	—	·9998	—	·9997	—	22·75
23·00					·9999⁹	—	·9999⁸	—	·9999⁷	—	·9999³	—	·9999⁻	—	·9998	—	23·00
23·25					1·0000⁻	—	·9999⁹	—	·9999⁸	—	·9999⁵	—	·9999¹	—	·9998	—	23·25
23·50							·9999⁹	—	·9999⁸	—	·9999⁸	—	·9999³	—	·9999⁻	—	23·50
23·75							·9999⁹	—	·9999⁹	—	·9999⁷	—	·9999⁵	—	·9999⁰	—	23·75
24·00							1·0000⁻	—	·9999⁹	—	·9999⁸	—	·9999⁶	—	·9999²	—	24·00
24·25									·9999⁹	—	·9999⁹	—	·9999⁷	—	·9999⁴	—	24·25
24·50									1·0000⁻	—	·9999⁹	—	·9999⁸	—	·9999⁶	—	24·50
24·75											·9999⁹	—	·9999⁸	—	·9999⁷	—	24·75
25·00											·9999⁹	—	·9999⁹	—	·9999⁸	—	25·00
25·25											1·0000⁻	—	·9999⁹	—	·9999⁸	—	25·25
25·50													·9999⁹	—	·9999⁹	—	25·50
25·75													1·0000⁻	—	·9999⁹	—	25·75
26·00															·9999⁹	—	26·00
26·25															·9999⁹	—	26·25
26·50															1·0000⁻	—	26·50

ON CORRECTIONS FOR THE MOMENT COEFFICIENTS OF FREQUENCY DISTRIBUTIONS WHEN THE START OF THE FREQUENCY IS ONE OF THE CHARACTERISTICS TO BE DETERMINED.

By E. S. MARTIN, B.Sc.

1. *Introduction.* In certain statistical data, for example, in cases of disease incidence in infancy, wages, incomes and house-valuations, the exact point at which the frequency commences is not evident. Thus the number of cases of Whooping Cough under one year of age may be recorded, but it is very unlikely that any of these occurred under three months; the number of houses valued for income-tax purposes under £20 per annum is published, but no house is actually valued at nothing per annum; the number of divorces occurring in the first year of marriage has been given, but it is clear that there is a certain minimum time which the legal formalities must take. Methods hitherto devised for fitting curves to such distributions have assumed the frequency to start at "zero," but this assumption may not only give poor results for the moments, but, in the case when the first frequency is less than the second, may make the distribution appear

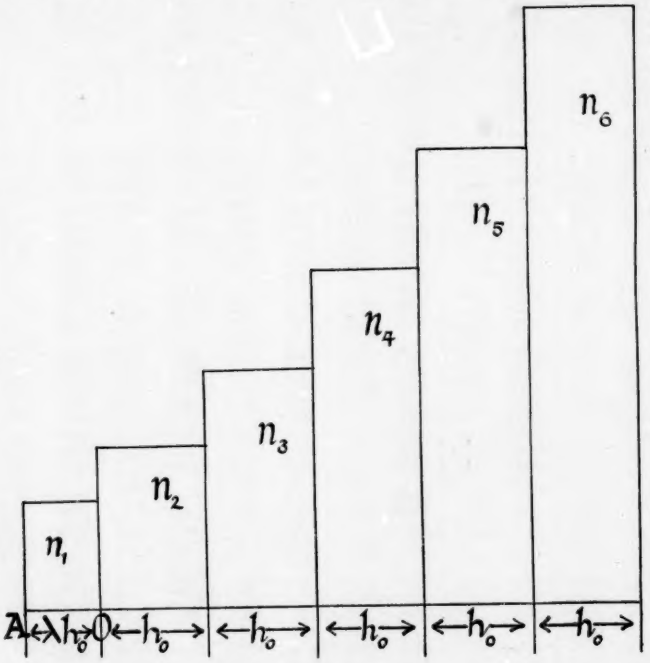


Fig. 1.

non-asymptotic when it is really asymptotic. The object of this paper is to investigate the possibility of determining the start of the frequency by means of an auxiliary curve and to examine what improvement in the evaluated moments of the distribution is thereby obtained.

2. Let the histogram of the first few sub-frequencies be as in Fig. 1, where the range of the first is λh_0 and the ranges of the others are h_0. Let there be p sub-ranges taken from $x=0$ (at A) to $x=x_p$. As in the paper by Eleanor Pairman and Karl Pearson*, let $Z=\int_x^{x_p} y\,dx$, and assume that Z is given in the neighbourhood of A by the 5th order parabola

$$Z=N+c_1\frac{x}{h_0}+c_2\left(\frac{x}{h_0}\right)^2+c_3\left(\frac{x}{h_0}\right)^3+c_4\left(\frac{x}{h_0}\right)^4+c_5\left(\frac{x}{h_0}\right)^5,$$

where N is the total frequency†. The c's are now determined in terms of λ by fitting the parabola to the first five frequencies. When

$$x=\lambda h_0, \qquad Z=N-n_1,$$
$$x=(1+\lambda)\,h_0, \quad Z=N-n_1-n_2, \text{ etc.,}$$

whence

$$\left.\begin{aligned}
-n_1 &= \lambda c_1+\lambda^2 c_2+\lambda^3 c_3+\lambda^4 c_4+\lambda^5 c_5\\
-n_1-n_2 &= (1+\lambda)\,c_1+(1+\lambda)^2 c_2+(1+\lambda)^3 c_3+(1+\lambda)^4 c_4+(1+\lambda)^5 c_5\\
-n_1-n_2-n_3 &= (2+\lambda)\,c_1+(2+\lambda)^2 c_2+(2+\lambda)^3 c_3+(2+\lambda)^4 c_4+(2+\lambda)^5 c_5\\
-n_1-n_2-n_3-n_4 &= (3+\lambda)\,c_1+(3+\lambda)^2 c_2+(3+\lambda)^3 c_3+(3+\lambda)^4 c_4+(3+\lambda)^5 c_5\\
-n_1-n_2-n_3-n_4-n_5 &= (4+\lambda)\,c_1+(4+\lambda)^2 c_2+(4+\lambda)^3 c_3+(4+\lambda)^4 c_4+(4+\lambda)^5 c_5
\end{aligned}\right\} \quad \text{.....(I).}$$

If we put

$$\lambda=\mu-2, \quad n_1=m_1, \quad n_1+n_2=m_2, \quad n_1+n_2+n_3=m_3,$$
$$n_1+n_2+n_3+n_4=m_4, \quad n_1+n_2+n_3+n_4+n_5=m_5,$$

the solution of these equations is

$$24c_1=-\frac{\mu(\mu^2-1)(\mu+2)}{\mu-2}m_1+4\frac{\mu(\mu+1)(\mu^2-4)}{\mu-1}m_2$$
$$-6\frac{(\mu^2-1)(\mu^2-4)}{\mu}m_3+4\frac{\mu(\mu-1)(\mu^2-4)}{\mu+1}m_4-\frac{\mu(\mu^2-1)(\mu-2)}{\mu+2}m_5,$$

$$24c_2=+\frac{(4\mu+2)(\mu^2+\mu-1)}{\mu-2}m_1-4\frac{4\mu^3+3\mu^2-8\mu-4}{\mu-1}m_2$$
$$+6\frac{4\mu^3-10\mu}{\mu}m_3-4\frac{4\mu^3-3\mu^2-8\mu+4}{\mu+1}m_4+\frac{(4\mu-2)(\mu^2-\mu-1)}{\mu+2}m_5,$$

* *Biometrika*, Vol. XII. pp. 231—258.

† In the paper by Pairman and Pearson the form

$$Z=N\left[1+c_1\left(\frac{x}{h_0}\right)+c_2\left(\frac{x}{h_0}\right)^2+c_3\left(\frac{x}{h_0}\right)^3+c_4\left(\frac{x}{h_0}\right)^4+c_5\left(\frac{x}{h_0}\right)^5\right]$$

was taken, but the above form is chosen here as it dispenses altogether with the use of "proportional frequencies."

$$24c_3 = -\frac{6\mu^2+6\mu-1}{\mu-2}m_1 + 4\frac{6\mu^2+3\mu-4}{\mu-1}m_2 - 6\frac{6\mu^2-5}{\mu}m_3 + 4\frac{6\mu^2-3\mu-4}{\mu+1}m_4 - \frac{6\mu^2-6\mu-1}{\mu+2}m_5,$$

$$24c_4 = +\frac{4\mu+2}{\mu-2}m_1 - 4\frac{4\mu+1}{\mu-1}m_2 + 6\frac{4\mu}{\mu}m_3 - 4\frac{4\mu-1}{\mu+1}m_4 + \frac{4\mu-2}{\mu+2}m_5,$$

$$24c_5 = -\frac{1}{\mu-2}m_1 + \frac{4}{\mu-1}m_2 - \frac{6}{\mu}m_3 + \frac{4}{\mu+1}m_4 - \frac{1}{\mu+2}m_5.$$

The coefficients in these equations have been evaluated for values of λ between 0 and 1. On substituting back for the m's in terms of the n's, the c's are obtained as linear functions of n_1, n_2, n_3, n_4, n_5. The coefficients in these expressions are given to twelve places of decimals in Table III.

Now let $n_1\mu_s''$ be the sth moment of the first frequency about O, i.e., about $x = \lambda h_0$, as given by the auxiliary curve. Then

$$n_1\mu_s'' = \int_0^{\lambda h_0} y\,(x-\lambda h_0)^s\,dx,$$

whence the first four moments are

$$n_1\mu_1'' = h_0\,(\tfrac{1}{2}\lambda^2 c_1 + \tfrac{1}{3}\lambda^3 c_2 + \tfrac{1}{4}\lambda^4 c_3 + \tfrac{1}{5}\lambda^5 c_4 + \tfrac{1}{6}\lambda^6 c_5),$$

$$n_1\mu_2'' = -h_0^2(\tfrac{1}{3}\lambda^3 c_1 + \tfrac{1}{6}\lambda^4 c_2 + \tfrac{1}{10}\lambda^5 c_3 + \tfrac{1}{15}\lambda^6 c_4 + \tfrac{1}{21}\lambda^7 c_5),$$

$$n_1\mu_3'' = h_0^3(\tfrac{1}{4}\lambda^4 c_1 + \tfrac{1}{10}\lambda^5 c_2 + \tfrac{1}{20}\lambda^6 c_3 + \tfrac{1}{35}\lambda^7 c_4 + \tfrac{1}{56}\lambda^8 c_5),$$

$$n_1\mu_4'' = -h_0^4(\tfrac{1}{5}\lambda^5 c_1 + \tfrac{1}{15}\lambda^6 c_2 + \tfrac{1}{35}\lambda^7 c_3 + \tfrac{1}{70}\lambda^8 c_4 + \tfrac{1}{126}\lambda^9 c_5).$$

With the help of Table III, these may be expressed as linear functions of n_1, n_2, n_3, n_4, n_5; the coefficients are given in Table IV.

Next, the formula

$$a_s = h_0^s\left(\frac{d^s Z}{dx^s}\right)_{x=\lambda h_0}$$

gives the abruptness coefficients at $x = \lambda h_0$ as linear functions of the *true* frequencies. These are

$$a_1 = c_1 + 2\lambda c_2 + 3\lambda^2 c_3 + 4\lambda^3 c_4 + 5\lambda^4 c_5,$$

$$a_2 = 2c_2 + 6\lambda c_3 + 12\lambda^2 c_4 + 20\lambda^3 c_5,$$

$$a_3 = 6c_3 + 24\lambda c_4 + 60\lambda^2 c_5,$$

$$a_4 = 24c_4 + 120\lambda c_5,$$

$$a_5 = 120c_5,$$

and again using Table III, the abruptness coefficients were calculated and are given in Table V.

Now suppose the raw moments of the remainder of the frequency, when n_1 is excluded, are calculated about O, using a sub-range h. Let them be $(N-n_1)\nu_1'''$,

etc., and write $\rho_0 = h/h_0$ and $a_s' = a_s \rho_0^s$. Then the corrected moments of the distribution about $x = \lambda h_0$ are as follows:

$$\left.\begin{aligned}
N\mu_1' &= (N-n_1)\,\nu_1''' + \tfrac{1}{12}h\,(a_1' - \tfrac{1}{60}a_3' + \tfrac{1}{2520}a_5') + n_1\mu_1'' \\
N\mu_2' &= (N-n_1)\,(\nu_2''' - \tfrac{1}{12}h^2) - \tfrac{1}{120}h^2\,(a_2' - \tfrac{5}{126}a_4') + n_1\mu_2'' \\
N\mu_3' &= (N-n_1)\,(\nu_3''' - \tfrac{1}{4}h^2\nu_1''') - \tfrac{1}{40}h^3\,(a_1' - \tfrac{5}{63}a_3' + \tfrac{1}{240}a_5') + n_1\mu_3'' \\
N\mu_4' &= (N-n_1)\,(\nu_4''' - \tfrac{1}{2}h^2\nu_2''' + \tfrac{7}{240}h^4) + \tfrac{1}{126}h^4(a_2' - \tfrac{7}{80}a_4') + n_1\mu_4''
\end{aligned}\right\} \quad \text{...(II),}$$

provided the distribution be not abrupt at the other end; in that case the usual corrections for abruptness must be applied for that terminal.

If $h = h_0$, then $a_s' = a_s$, and the equations may be simplified by putting

$$K_1 = \tfrac{1}{12}\,(a_1 - \tfrac{1}{60}a_3 + \tfrac{1}{2520}a_5) + \frac{1}{h}\,n_1\mu_1'',$$

$$K_2 = -\tfrac{1}{120}\,(a_2 - \tfrac{5}{126}a_4) + \frac{1}{h^2}\,n_1\mu_2'',$$

$$K_3 = -\tfrac{1}{40}\,(a_1 - \tfrac{5}{63}a_3 + \tfrac{1}{240}a_5) + \frac{1}{h^3}\,n_1\mu_3'',$$

$$K_4 = \tfrac{1}{126}\,(a_2 - \tfrac{7}{80}a_4) + \frac{1}{h^4}\,n_1\mu_4'',$$

whence the corrected moments are

$$\left.\begin{aligned}
N\mu_1' &= (N-n_1)\,\nu_1''' + hK_1 \\
N\mu_2' &= (N-n_1)\,(\nu_2''' - \tfrac{1}{12}h^2) + h^2K_2 \\
N\mu_3' &= (N-n_1)\,(\nu_3''' - \tfrac{1}{4}h^2\nu_1''') + h^3K_3 \\
N\mu_4' &= (N-n_1)\,(\nu_4''' - \tfrac{1}{2}h^2\nu_2''' + \tfrac{7}{240}h^4) + h^4K_4
\end{aligned}\right\} \quad \text{.....(III).}$$

The values of the K's are given in Table VI.

3. We now come to the most troublesome part of the problem: namely, to determine from a given set of frequencies the best value of λ. Three methods will be proposed.

1st Method. Following on the lines of Miss Pearse's paper on asymptotic distributions*, we may select λ by making our parabola fit the sixth frequency n_6. When $x = (5+\lambda)\,h_0$, $Z = N - n_1 - n_2 - n_3 - n_4 - n_5 - n_6$ and on substituting in the equation of the parabola we get

$$n_6 = -n_1 - n_2 - n_3 - n_4 - n_5 - (5+\lambda)\,c_1 - (5+\lambda)^2\,c_2 - (5+\lambda)^3\,c_3 - (5+\lambda)^4\,c_4 - (5+\lambda)^5\,c_5,$$

giving n_6 as a linear function of n_1, n_2, n_3, n_4, n_5, whose coefficients are given in Table II.

In order to select λ in any given case, the values of n_6 are calculated from this table for various values of λ and that value of λ is chosen which yields a result nearest to the given n_6.

* "On Corrections for the Moment Coefficients of Frequency Distributions when there are Infinite Ordinates at one or both of the Terminals of the Range," *Biometrika*, Vol. xx^A. pp. 314—355.

This method does not always work very well, thus (*a*) the value of λ so found sometimes differs widely from the true value, (*b*) the table may not give any value of λ, i.e., there is no auxiliary parabola, for which λ is less than unity, which fits the first six frequencies, (*c*) the table may give two values of λ, i.e., there are two such parabolas.

2nd Method. If it is expected that the final frequency curve will rise abruptly from the x-axis, we may, instead of making our parabola fit the sixth frequency, impose the condition that its initial ordinate shall be zero. Now since $y = -\frac{dZ}{dx}$, the ordinate at $x = 0$ is $-\frac{c_1}{h_0}$. Hence we may use Table III to find that value of λ which makes c_1 most nearly equal to zero. This test, while not always giving an accurate result, has been found in several cases to give a better result than the first test, and it has not yet in any case been found to give a double value of λ.

3rd Method. The poor results which the above tests sometimes give are no doubt due to the inadequacy of our parabola to represent various types of curves.

Accordingly. for the purpose of choosing λ a new type of auxiliary curve may be proposed.

(i) If $Z = N + Ax^q$ be taken as auxiliary curve*, then Z contains three unknowns, A, q and λ. We may hence fit this curve to the first three frequencies. We have

$$-n_1 = A\lambda^q,$$
$$-n_1 - n_2 = A(1+\lambda)^q,$$
$$-n_1 - n_2 - n_3 = A(2+\lambda)^q.$$

Writing $n_1 = m_1, \quad n_1 + n_2 = m_2, \quad n_1 + n_2 + n_3 = m_3,$

we get, eliminating A,

$$\frac{m_2}{m_1} = \left(\frac{1+\lambda}{\lambda}\right)^q, \quad \frac{m_3}{m_2} = \left(\frac{2+\lambda}{1+\lambda}\right)^q,$$

whence

$$\frac{\log\left(\frac{m_2}{m_1}\right)}{\log\left(\frac{m_3}{m_2}\right)} = \frac{\log\left(\frac{1+\lambda}{\lambda}\right)}{\log\left(\frac{2+\lambda}{1+\lambda}\right)}.$$

is the equation from which to determine λ. Writing

$$\frac{\log\left(\frac{m_2}{m_1}\right)}{\log\left(\frac{m_3}{m_2}\right)} = \frac{{}_3M_1}{{}_3M_2} \text{ and } \frac{\log\left(\frac{1+\lambda}{\lambda}\right)}{\log\left(\frac{2+\lambda}{1+\lambda}\right)} = \frac{{}_3L_1}{{}_3L_2},$$

we have calculated the values of the latter ratio for various values of λ, the results being given in Table I. All the computer has to do, therefore, is to calculate the value of $\frac{{}_3M_1}{{}_3M_2}$ from the given frequencies and, referring to Table I, select that value of λ for which $\frac{{}_3L_1}{{}_3L_2}$ is most nearly equal to the computed value of $\frac{{}_3M_1}{{}_3M_2}$.

* x stands here for x/h_0.

If q also be required, we have $q = \frac{{}_3M_1}{{}_3L_1}$ and the appropriate value of ${}_3L_1$ can be taken from the column adjacent to $\frac{{}_3L_1}{{}_3L_2}$.

(ii) Now three frequencies will often be insufficient from which to determine λ, especially if h_0 be large. To get a better result we may fit the curve

$$Z = N + Ax^q e^{ax}$$

to the first four frequencies. Eliminating A, q and a from the resulting equations by successive divisions we get

$$\frac{\log\left(\frac{m_2^2}{m_1 m_3}\right)}{\log\left(\frac{m_3^2}{m_2 m_4}\right)} = \frac{\log\left\{\frac{(1+\lambda)^2}{\lambda(2+\lambda)}\right\}}{\log\left\{\frac{(2+\lambda)^2}{(1+\lambda)(3+\lambda)}\right\}},$$

or
$$\frac{{}_4M_1}{{}_4M_2} = \frac{{}_4L_1}{{}_4L_2},$$

and
$$q = \frac{{}_4M_1}{{}_4L_1}.$$

The values of $\frac{{}_4L_1}{{}_4L_2}$ and ${}_4L_1$ are given in Table I and the procedure is as for three frequencies.

(iii) Similarly, fitting $Z = N + Ax^q e^{ax+bx^2}$

to the first five frequencies, we have

$$\frac{\log\left(\frac{m_2^3 m_4}{m_1 m_3^3}\right)}{\log\left(\frac{m_3^3 m_5}{m_2 m_4^3}\right)} = \frac{\log\left\{\frac{(1+\lambda)^3(3+\lambda)}{\lambda(2+\lambda)^3}\right\}}{\log\left\{\frac{(2+\lambda)^3(4+\lambda)}{(1+\lambda)(3+\lambda)^3}\right\}},$$

or
$$\frac{{}_5M_1}{{}_5M_2} = \frac{{}_5L_1}{{}_5L_2},$$

and
$$q = \frac{{}_5M_1}{{}_5L_1}.$$

The values of $\frac{{}_5L_1}{{}_5L_2}$ and ${}_5L_1$ are given in Table I.

(iv) If the curve $Z = N + Ax^q e^{ax+bx^2+cx^3}$

be fitted to the first six frequencies, the equation for λ is

$$\frac{\log\left(\frac{m_2^4 m_4^4}{m_1 m_3^6 m_5}\right)}{\log\left(\frac{m_3^4 m_5^4}{m_2 m_4^6 m_6}\right)} = \frac{\log\left\{\frac{(1+\lambda)^4(3+\lambda)^4}{\lambda(2+\lambda)^6(4+\lambda)}\right\}}{\log\left\{\frac{(2+\lambda)^4(4+\lambda)^4}{(1+\lambda)(3+\lambda)^6(5+\lambda)}\right\}},$$

or
$$\frac{{}_6M_1}{{}_6M_2} = \frac{{}_6L_1}{{}_6L_2},$$

and
$$q = \frac{{}_6M_1}{{}_6L_1}.$$

The values of $\frac{{}_6L_1}{{}_6L_2}$ and ${}_6L_1$ are given in Table I.

(v) If the curve $$Z = N + Ax^q e^{ax+bx^2+cx^3+dx^4}$$

be fitted to the first seven frequencies, the equation for λ is

$$\frac{\log\left(\frac{m_2^5 m_4^{10} m_6}{m_1 m_3^{10} m_5^5}\right)}{\log\left(\frac{m_3^5 m_5^{10} m_7}{m_2 m_4^{10} m_6^5}\right)} = \frac{\log\left\{\frac{(1+\lambda)^5(3+\lambda)^{10}(5+\lambda)}{\lambda(2+\lambda)^{10}(4+\lambda)^5}\right\}}{\log\left\{\frac{(2+\lambda)^5(4+\lambda)^{10}(6+\lambda)}{(1+\lambda)(3+\lambda)^{10}(5+\lambda)^5}\right\}},$$

or $$\frac{{}_7M_1}{{}_7M_2} = \frac{{}_7L_1}{{}_7L_2},$$

and $$q = \frac{{}_7M_1}{{}_7L_2}.$$

The values of $\frac{{}_7L_1}{{}_7L_2}$ and ${}_7L_1$ are given in Table I.

This method, though apparently rather elaborate, is simple in application. If the logarithms of m_1, m_2, m_3, m_4, m_5, m_6 and m_7 are found, the five values of λ corresponding to 3, 4, 5, 6 and 7 frequencies are quickly determined*. In any theoretical distribution it will be found that the more frequencies are taken, the more nearly does the corresponding value of λ approach the true value. In fact, among numerous theoretical frequencies that have been tested, we have not found one in which the test fails to yield the correct value of λ (to one significant figure) when the seven initial frequencies are taken. In experimental frequencies, however, this is not the case, as the values of M_1 and M_2 become more and more affected by the sampling errors as more frequencies are taken. Thus the value of λ given by seven frequencies may be an impossible one, while that given by four or five frequencies may yield good results. For this reason the present test may not be as efficacious as the previous tests in any particular case, but it has one advantage over them; it can be used for either asymptotic or non-asymptotic frequencies. We have placed no limit on the value of q and if it is found that $0 < q < 1$ then an asymptotic frequency is indicated. It may be remarked, however, that although the auxiliary parabola of the n_6 test is incapable of having an infinite initial ordinate, yet it can have a very large one and will yield reasonable results when applied to an asymptotic frequency, provided the degree of asymptoticity is not large.

Our third method will be referred to as the "e test."

As an example consider the Type VI curve

$$y = 33 \times (160)^6 \frac{x-4}{x^{13}}, \text{ starting at } x = 4.$$

Let $h_0 = \cdot 2$, $\lambda = \cdot 5$, then from the formula

$$\int_4^x y dx = 10^6 \left(1 - \frac{12}{(x/4)^{11}} + \frac{11}{(x/4)^{12}}\right)$$

* We note that: ${}_7M_1 = -\log m_1 + 5\log m_2 - 10\log m_3 + 10\log m_4 - 5\log m_5 + \log m_6$, and similarly for the other M's.

we have the "frequencies

$$n_1 = 33{,}377, \qquad n_5 = 123{,}600,$$
$$n_2 = 168{,}900, \qquad n_6 = 91{,}023,$$
$$n_3 = 189{,}434, \qquad n_7 = 65{,}653.$$
$$n_4 = 160{,}671,$$

Tables II and III give the following results:

λ	n_6	c_1
·1	966,781	− 368,500
·2	276,218	− 171,696
·3	111,069	− 90,182
·4	61,994	− 39,885
·5	51,108	− 3,641
·6	54,606	+ 24,479
·7	63,513	+ 47,148
·8	74,072	+ 65,788
·9	84,638	+ 81,248
1·0	94,507	+ 94,078

These results are shown in Figs. 2 and 3. It is seen that the n_6 test gives (without interpolation) the two values ·3 and 1·0 for λ, while the c_1 test gives the correct value ·5.

Applying our third test we have

$$m_1 = 33{,}377, \qquad \log m_1 = 4{\cdot}523\,4473,$$
$$m_2 = 202{,}277, \qquad \log m_2 = 5{\cdot}305\,9465,$$
$$m_3 = 391{,}711, \qquad \log m_3 = 5{\cdot}592\,9658,$$
$$m_4 = 552{,}382, \qquad \log m_4 = 5{\cdot}742\,2395,$$
$$m_5 = 675{,}982, \qquad \log m_5 = 5{\cdot}829\,9351,$$
$$m_6 = 767{,}005, \qquad \log m_6 = 5{\cdot}884\,7983,$$
$$m_7 = 832{,}658, \qquad \log m_7 = 5{\cdot}920\,4667.$$

Taking in each case the nearest value of λ (without interpolation) from Table I, we get the following results:

3 frequencies. $\dfrac{{}_3M_1}{{}_3M_2} = \dfrac{\cdot 782\,4992}{\cdot 287\,0193} = 2{\cdot}7263, \qquad \lambda = \cdot 3.$

4 frequencies. $\dfrac{{}_4M_1}{{}_4M_2} = \dfrac{\cdot 495\,4799}{\cdot 137\,7456} = 3{\cdot}5971, \qquad \lambda = \cdot 4.$

5 frequencies. $\dfrac{{}_5M_1}{{}_5M_2} = \dfrac{\cdot 357\,7343}{\cdot 076\,1675} = 4{\cdot}6967, \qquad \lambda = \cdot 5.$

6 frequencies. $\dfrac{{}_6M_1}{{}_6M_2} = \dfrac{\cdot 281\,5668}{\cdot 047\,4218} = 5{\cdot}9375, \qquad \lambda = \cdot 5.$

7 frequencies. $\dfrac{{}_7M_1}{{}_7M_2} = \dfrac{\cdot 234\,1450}{\cdot 032\,3137} = 7{\cdot}246, \qquad \lambda = \cdot 5.$

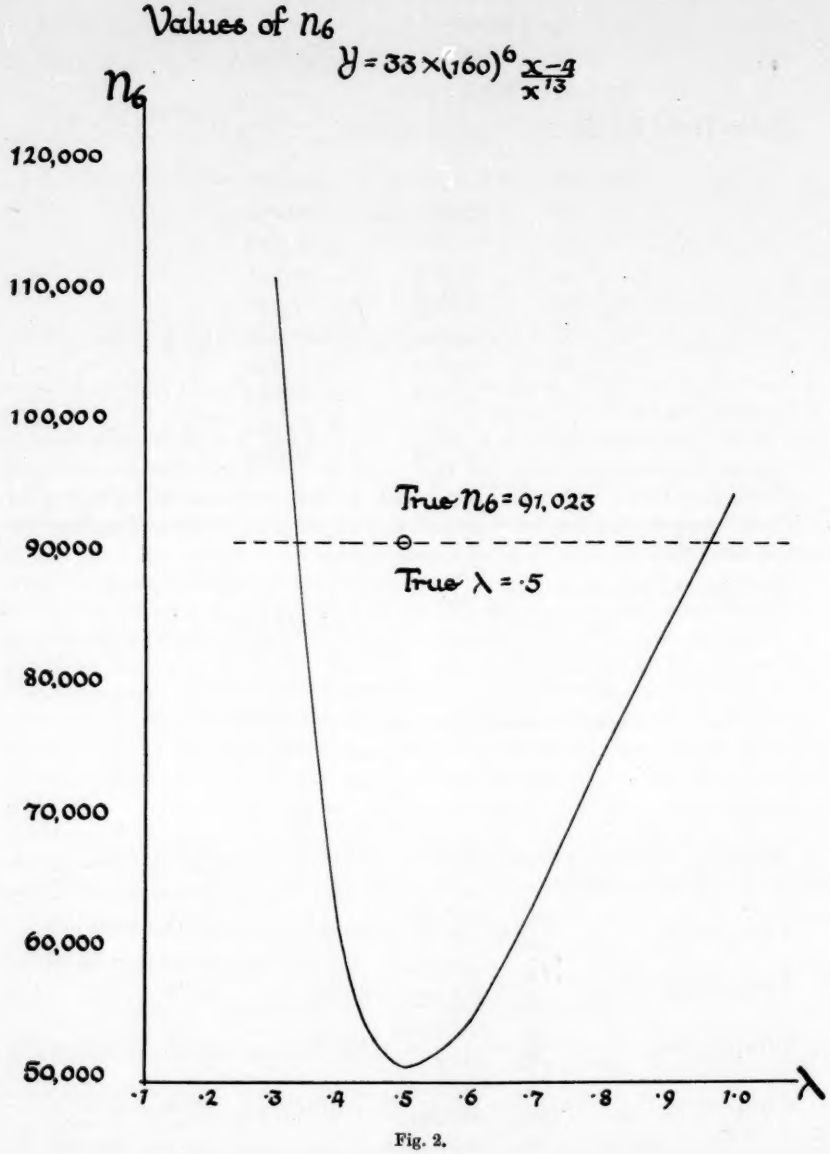

Values of n_6
$y = 33 \times (160)^6 \frac{x-4}{x^{13}}$
n_6
120,000
110,000
100,000
90,000
80,000
70,000
60,000
50,000
True n_6 = 91,023
True λ = ·5
·1
·2
·3
·4
·5
·6
·7
·8
·9
1·0
λ

Fig. 2.

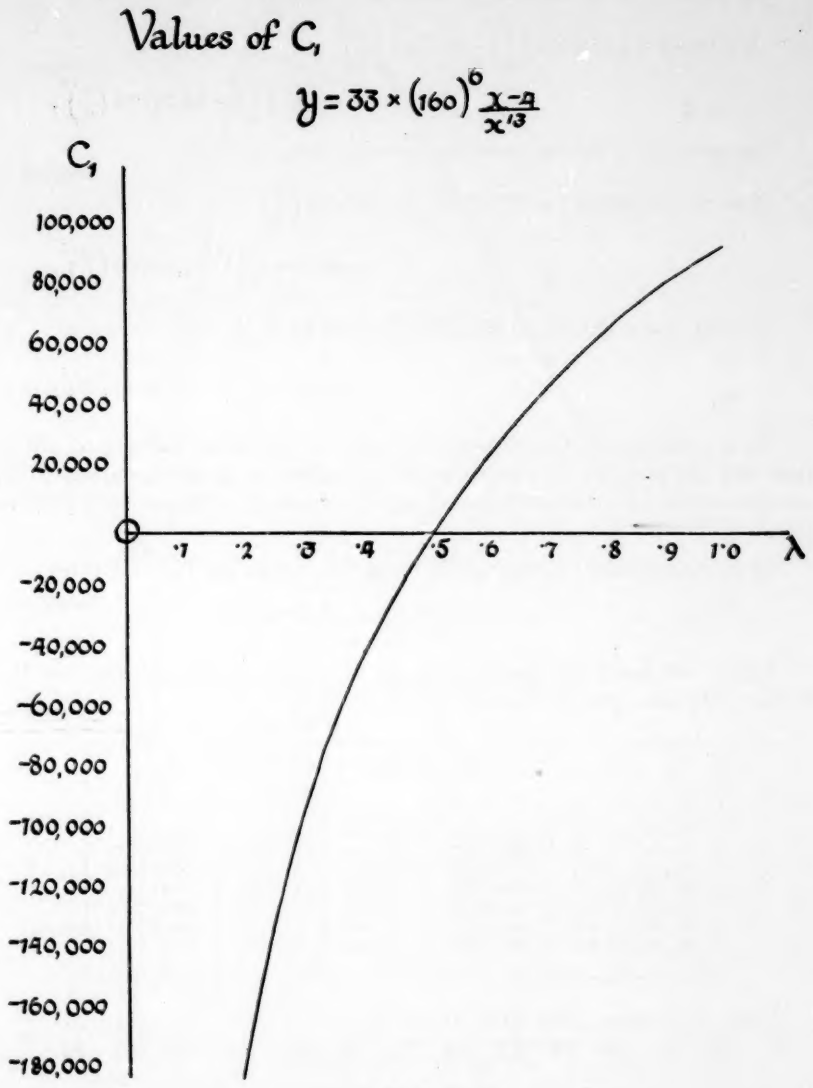

Values of C,
$y = 33 \times (160)^6 \frac{x-9}{x^{13}}$
C,
100,000
80,000
60,000
40,000
20,000
-20,000
-40,000
-60,000
-80,000
-100,000
-120,000
-140,000
-160,000
-180,000
·1 ·2 ·3 ·4 ·5 ·6 ·7 ·8 ·9 1·0 λ

Fig. 3.

Hence the correct value of λ is obtained by the use of five frequencies.

The equation of the auxiliary parabola with the correct value ·5 for λ is

$$y = 18{,}206 + 1{,}513{,}518\left(\frac{x}{\cdot 2}\right) - 817{,}671\left(\frac{x}{\cdot 2}\right)^2 + 176{,}037\left(\frac{x}{\cdot 2}\right)^3 - 14{,}187{\cdot}94\left(\frac{x}{\cdot 2}\right)^4.$$

The equations of the two parabolas given by the n_6 test are

$$\lambda = \cdot 3,\quad y = 450{,}911 + 771{,}550\left(\frac{x}{\cdot 2}\right) - 356{,}436\left(\frac{x}{\cdot 2}\right)^2 + 49{,}760{\cdot}9\left(\frac{x}{\cdot 2}\right)^3 - 1{,}696{\cdot}00\left(\frac{x}{\cdot 2}\right)^4,$$

$$\lambda = 1{\cdot}0,\quad y = -470{,}389 + 1{,}656{,}578\left(\frac{x}{\cdot 2}\right) - 641{,}893\left(\frac{x}{\cdot 2}\right)^2 + 95{,}914{\cdot}6\left(\frac{x}{\cdot 2}\right)^3 - 5{,}146{\cdot}46\left(\frac{x}{\cdot 2}\right)^4.$$

These parabolas together with the original curve are shown in Fig. 4. It will be seen that the parabola for which $\lambda = 1{\cdot}0$ has, besides its incorrect starting-point, another serious fault; its first frequency n_1 is composed of a negative and a positive part.

As an example of an asymptotic frequency let us take the Type III curve

$$y = \frac{10^6}{\Gamma(\cdot 9)}\, x^{-\cdot 1} e^{-x}.$$

Take $\lambda = \cdot 8$, $h_0 = \frac{1}{2}\sqrt{\cdot 9}$, then $\lambda h_0 = \cdot 4\sqrt{\cdot 9}$. The first seven frequencies, from the *Tables of the Incomplete* Γ*-Function*, are shown below:

x	u	$I(u, p)$	n	m	$\log m$
$\cdot 4\sqrt{\cdot 9}$	·4	·365 4305	365,431	365,431	5·562 8054
$\cdot 9\sqrt{\cdot 9}$	·9	·620 7737	255,343	620,774	5·792 9335
$1{\cdot}4\sqrt{\cdot 9}$	1·4	·770 3535	149,580	770,354	5·886 6903
$1{\cdot}9\sqrt{\cdot 9}$	1·9	·860 0470	89,693	860,047	5·934 5222
$2{\cdot}4\sqrt{\cdot 9}$	2·4	·914 3798	54,333	914,380	5·961 1267
$2{\cdot}9\sqrt{\cdot 9}$	2·9	·947 4827	33,103	947,483	5·976 5715
$3{\cdot}4\sqrt{\cdot 9}$	3·4	·967 7261	20,243	967,726	5·985 7524

Using the n column with Table II, we have

$$\lambda = \cdot 8,\quad n_6 = 88{,}314;$$
$$\lambda = \cdot 9,\quad n_6 = 9{,}583.$$

Hence, since n_6 is really 33,103, the n_6 test gives ·9 as the nearest value of λ, or ·87 by linear interpolation.

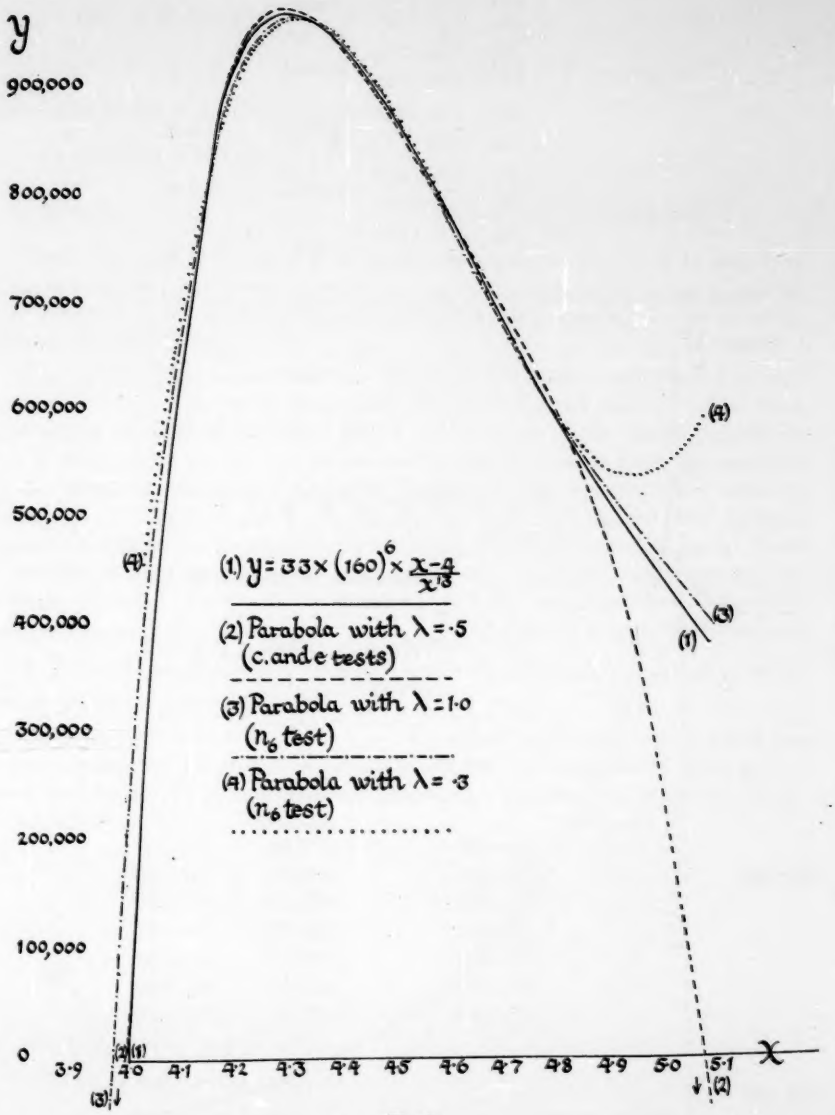

y
900,000
800,000
700,000
600,000
500,000
400,000
300,000
200,000
100,000
0
3·9
4·0
4·1
4·2
4·3
4·4
4·5
4·6
4·7
4·8
4·9
5·0
5·1
x
(1) $y = 33 \times (160)^6 \times \frac{x-4}{x^{13}}$
(2) Parabola with λ = ·5 (c. and e tests)
(3) Parabola with λ = 1·0 (n_6 test)
(4) Parabola with λ = ·3 (n_6 test)

Fig. 4.

Using the log m column, we get

$$\frac{{}_3M_1}{{}_3M_2} = \frac{\cdot 230\,1281}{\cdot 093\,7568} = 2\cdot 4545, \qquad \lambda = \cdot 3.$$

$$\frac{{}_4M_1}{{}_4M_2} = \frac{\cdot 136\,3713}{\cdot 045\,9249} = 2\cdot 9694, \qquad \lambda = \cdot 7.$$

$$\frac{{}_5M_1}{{}_5M_2} = \frac{\cdot 090\,4464}{\cdot 024\,6975} = 3\cdot 6622, \qquad \lambda = \cdot 8.$$

$$\frac{{}_6M_1}{{}_6M_2} = \frac{\cdot 065\,7489}{\cdot 014\,6298} = 4\cdot 4942, \qquad \lambda = \cdot 8.$$

$$\frac{{}_7M_1}{{}_7M_2} = \frac{\cdot 051\,1191}{\cdot 009\,4579} = 5\cdot 405, \qquad \lambda = \cdot 8.$$

Hence the correct value of λ is again obtained by the use of five frequencies.

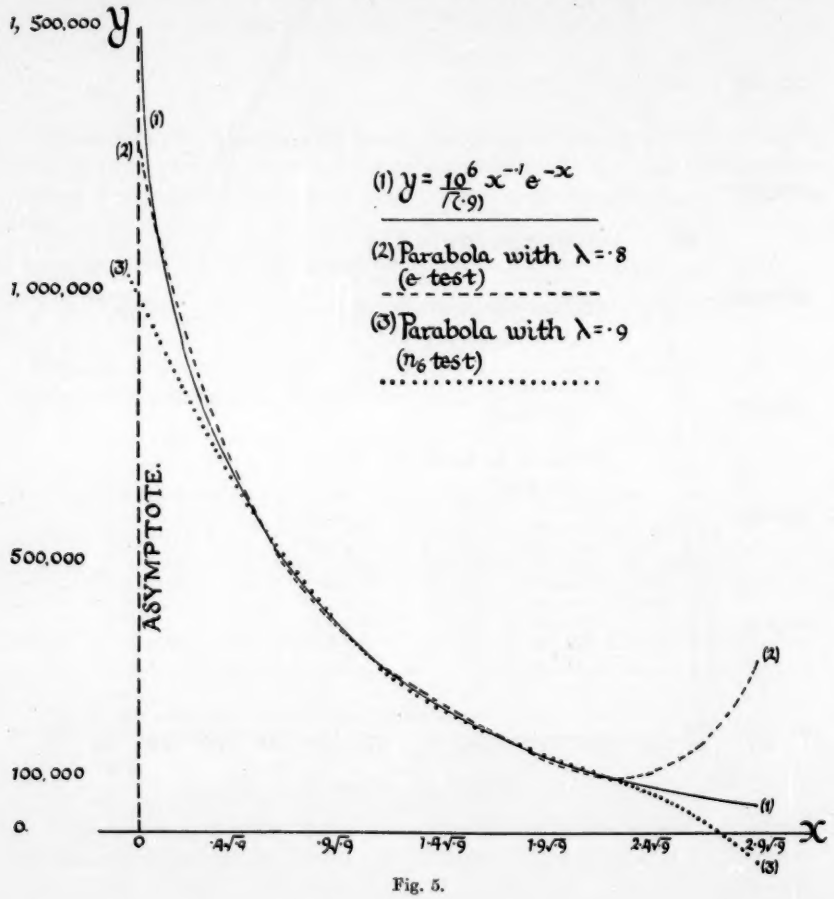


Fig. 5.

The equation of the parabola with $\lambda = \cdot 8$, referred to its start, is

$$y = 1{,}263{,}615 - 919{,}415\left(\frac{x}{h_0}\right) + 356{,}072\left(\frac{x}{h_0}\right)^2 - 72{,}594{\cdot}5\left(\frac{x}{h_0}\right)^3 + 5{,}811{\cdot}89\left(\frac{x}{h_0}\right)^4,$$

and of that with $\lambda = \cdot 9$, referred to its start, is

$$y = 1{,}038{,}998 - 438{,}479\left(\frac{x}{h_0}\right) + 49{,}970{\cdot}5\left(\frac{x}{h_0}\right)^2 + 5{,}305{\cdot}22\left(\frac{x}{h_0}\right)^3 - 1{,}104{\cdot}01\left(\frac{x}{h_0}\right)^4.$$

These, together with the original curve, are shown in Fig. 5. It is seen that there is very close agreement between the curves for frequencies n_2, n_3, n_4, n_5 and that there is a good attempt at agreement at the start, especially in the parabola with the correct value of λ.

The curves of our e-test would no doubt fit still better, and there is in fact some inconsistency in using these curves to determine the value of λ and then reverting to our original parabola. But it will be shown in the following examples that if the correct value of λ be determined, the corrections given to the moments by the parabola are usually adequate. Further, if the e-curves were used for correcting the moments, the K's would have to be worked out *ab initio* for each example, owing to the impracticability of tabling for values of both λ *and* q. There is the further objection that with these curves the abruptness coefficients tend to diverge. We shall therefore be content to correct the moments by the use of the tables at the end of this paper, no matter how the value of λ may be determined.

4. We will now investigate, in some further distributions, the extent to which the moments are corrected by the method of this paper.

Example 1. Consider the curve $y = 156 \times 10^6\, x(1-x)^{11}$ from 0 to 1, which has a total "frequency" 1,000,000. We will take our first sub-range as ·02 starting from $x = 0$ and the other sub-ranges as ·06; thus $\lambda = \frac{1}{3}$. The frequencies are as follows:

x	Frequency	x	Frequency
0—·02	26,951,	·38—·44	11,965,
·02—·08	252,423,	·44—·50	4,264,
·08—·14	281,978,	·50—·56	1,302,
·14—·20	205,002,	·56—·62	330,
·20—·26	122,555,	·62—·68	66,
·26—·32	63,781,	·68—·74	10,
·32—·38	29,372,	·74—·80	1.

We have first to find λ.

(i) The n_6 test. Table II gives:

when $\lambda = \cdot 2$, $n_6 = 76{,}551$; when $\lambda = \cdot 3$, $n_6 = -\ 2{,}153$;

when $\lambda = \cdot 6$, $n_6 = 53{,}936$; when $\lambda = \cdot 7$, $n_6 = \ 81{,}889$.

Since n_6 is really 63,781, this test gives the two values ·2 and ·6 for λ.

(ii) The c_1 test. Table III gives:

when $\lambda = \cdot 3$, $c_1 = -22{,}007$; when $\lambda = \cdot 4$, $c_1 = 30{,}348$.

Hence c_1 is most nearly equal to zero when $\lambda = \cdot 3$, or $\cdot 34$ by linear interpolation. This is a good result.

(iii) The e test. We have

$m_1 = 26{,}951$, $\log m_1 = 4{\cdot}430\ 5749$,
$m_2 = 279{,}374$, $\log m_2 = 5{\cdot}446\ 1860$,
$m_3 = 561{,}352$, $\log m_3 = 5{\cdot}749\ 2352$,
$m_4 = 766{,}354$, $\log m_4 = 5{\cdot}884\ 4295$,
$m_5 = 888{,}909$, $\log m_5 = 5{\cdot}948\ 8573$,
$m_6 = 952{,}690$, $\log m_6 = 5{\cdot}978\ 9516$,
$m_7 = 982{,}062$, $\log m_7 = 5{\cdot}992\ 1389$.

Whence, using Table I we have

$$\frac{{}_3M_1}{{}_3M_2} = 3{\cdot}35131, \quad \lambda = \cdot 15^-,$$

$$\frac{{}_4M_1}{{}_4M_2} = 4{\cdot}24511, \quad \lambda = \cdot 31,$$

$$\frac{{}_5M_1}{{}_5M_2} = 5{\cdot}6104, \quad \lambda = \cdot 35^+,$$

$$\frac{{}_6M_1}{{}_6M_2} = 7{\cdot}3797, \quad \lambda = \cdot 34,$$

$$\frac{{}_7M_1}{{}_7M_2} = 9{\cdot}291, \quad \lambda = \cdot 333.$$

The values of λ are obtained by linear interpolation. It is seen that 4 frequencies give a value correct to 1 significant figure while 7 frequencies give a very exact value.

Let us take $\lambda = \cdot 3$ and test Table IV to find the accuracy with which our auxiliary curve gives the moments of the first frequency about $x = \cdot 02$. We get the following results:

	From Table	True value
$n_1\mu_1''$	$-185{\cdot}36$	$-186{\cdot}44$
$n_1\mu_2''$	$+\ 1{\cdot}868$	$+\ 1{\cdot}906$
$n_1\mu_3''$	$-\ \cdot 022\ 25$	$-\ \cdot 023\ 21$
$n_1\mu_4''$	$+\ \cdot 000\ 2914$	$+\ \cdot 000\ 3126$

It is seen that the first moment is in error in the units figure, but since the total first moment of the whole curve about $x = \cdot 02$ is $122{,}857\frac{1}{7}$ this error will not seriously affect the results. Similarly the total second moment about $x = \cdot 02$ is $23{,}257\frac{1}{7}$, and hence the error in the second moment will not affect the sixth significant figure. Also since the total third and fourth moments are respectively 5592 and $1593{\cdot}572$, we should again get at least six figures correct on adding the

corresponding moments of the first frequency. We can therefore say that the moments of the first frequency as given by Table IV are sufficiently accurate.

We will now find the corrected moments of the total frequency about $x = \cdot 02$, taking $h = 1$ as our working sub-range. With the notation of Equations III (p. 15) above,

$$(N - n_1)\,\nu_1''' = 2{,}063{,}961{\cdot}5, \qquad (N - n_1)\,\nu_3''' = 26{,}400{,}593{\cdot}4,$$
$$(N - n_1)\,\nu_2''' = 6{,}538{,}164{\cdot}25, \qquad (N - n_1)\,\nu_4''' = 126{,}209{,}278{\cdot}6.$$

The values of the K's, obtained by using $\lambda = \cdot 3$ in Table VI, are

$$K_1 = -\,16{,}220{\cdot}0, \qquad K_3 = +\,4605{\cdot}9,$$
$$K_2 = +\,3257{\cdot}5, \qquad K_4 = -\,1955{\cdot}0.$$

The corrected moments are therefore

$$N\mu_1' = 2{,}063{,}961{\cdot}5 - 16{,}220{\cdot}0 = 2{,}047{,}742,$$
$$N\mu_2' = 6{,}538{,}164{\cdot}25 - 81{,}087{\cdot}42 + 3257{\cdot}5 = 6{,}460{,}334,$$
$$N\mu_3' = 26{,}400{,}593{\cdot}4 - 515{,}990{\cdot}4 + 4605{\cdot}9 = 25{,}889{,}209,$$
$$N\mu_4' = 126{,}209{,}278{\cdot}6 - 3{,}269{,}082{\cdot}1 + 28{,}380{\cdot}6 - 1955{\cdot}0 = 122{,}966{,}622,$$

whence

$$\mu_1' = 2{\cdot}047\,742, \qquad \mu_3' = 25{\cdot}889\,209,$$
$$\mu_2' = 6{\cdot}460\,334, \qquad \mu_4' = 122{\cdot}966\,622.$$

Transferring to the mean as here given, viz., 2·047 742 from the start of the second frequency, we obtain

$$\mu_2 = 2{\cdot}267\,087,$$
$$\mu_3 = 3{\cdot}375\,294,$$
$$\mu_4 = 20{\cdot}697\,620.$$

Now the true moments about the mean are easily found from the equation of the curve, and can be brought to the working scale by dividing μ_s by $(\cdot 06)^s$. They are shown in the following comparative table:

	By present method	True value	% error
Mean from start of second frequency	2·047 742	2·047 619	+ ·006
μ_2	2·267 087	2·267 574	− ·021
μ_3	3·375 294	3·374 366	+ ·028
μ_4	20·697 620	20·699 890	− ·011

It is seen that the percentage errors in all four moments are very small.

It is of interest to compare these results with those which would be obtained by supposing the first frequency group on the same base as the others and finding the moments about its supposed start. We shall use the abruptness coefficients as given in the paper by Pairman and Pearson cited above. The raw moment-coefficients are

$$\nu_1' = 3{\cdot}050\,486, \qquad \nu_3' = 53{\cdot}183\,388,$$
$$\nu_2' = 11{\cdot}645\,874, \qquad \nu_4' = 280{\cdot}271\,217.$$

These, if taken without correction, give the following results:

Mean from start of second frequency = 2·050 486, error = + ·140%,

$$\nu_2 = 2{\cdot}340\,409, \quad \text{error} = +3{\cdot}212\%,$$
$$\nu_3 = 3{\cdot}379\,081, \quad \text{error} = + \cdot140\%,$$
$$\nu_4 = 21{\cdot}776\,860, \quad \text{error} = +5{\cdot}203\%.$$

Applying Sheppard's corrections we get (ν_1' and ν_3 being uncorrected)

$$\nu_2' - \tfrac{1}{12} = 11{\cdot}562\,541,$$
$$\nu_3' - \tfrac{1}{4}\nu_1' = 52{\cdot}420\,767,$$
$$\nu_4' - \tfrac{1}{2}\nu_2' + \tfrac{7}{240} = 274{\cdot}477\,447.$$

With these corrections alone, we get

$$\mu_2 = 2{\cdot}257\,076, \quad \text{error} = -\cdot463\%,$$
$$\mu_4 = 20{\cdot}635\,988, \quad \text{error} = -\cdot309\%,$$

which are a good improvement on the raw moment results.

Now the usual abruptness corrections* work out to be

$$H_1 = -\cdot013\,828, \qquad H_3 = -\cdot003\,662,$$
$$H_2 = -\cdot004\,091, \qquad H_4 = -\cdot003\,872,$$

whence the fully corrected moment-coefficients are

$$\mu_1' = \nu_1' - H_1 = 3{\cdot}064\,314,$$
$$\mu_2' = \nu_2' - \tfrac{1}{12} - H_2 = 11{\cdot}566\,632,$$
$$\mu_3' = \nu_3' - \tfrac{1}{4}\nu_1' + H_3 = 52{\cdot}417\,105,$$
$$\mu_4' = \nu_4' - \tfrac{1}{2}\nu_2' + \tfrac{7}{240} + H_5 = 274{\cdot}473\,575.$$

Hence H_1 acts the wrong way and gives a worse value for the mean than the raw first moment:

Mean from start of second frequency = 2·064 314, error = + ·815%.

Referred to this mean,

$$\mu_2 = 2{\cdot}176\,612, \quad \text{error} = -4{\cdot}011\%,$$
$$\mu_3 = 3{\cdot}633\,669, \quad \text{error} = +7{\cdot}685\%,$$
$$\mu_4 = 19{\cdot}131\,702, \quad \text{error} = -7{\cdot}576\%.$$

These are obviously much worse than the raw moments alone.

Thus we have the remarkable result that Sheppard's corrections, which are obviously inapplicable since there is not high contact at the start, give very good results, while the usual abruptness corrections give very bad results. Now it is shown in the paper by Pairman and Pearson that the reverse is often the case, and

* F. Sandon, *Biometrika*, Vol. XVI. pp. 193—195 (with an error in the value for μ_2'), and *Tables for Statisticians and Biometricians*, Part II. p. cxciii, Introduction (error corrected).

hence one must conclude that the present result is due to the incorrect value given to the base of the first frequency. This in fact has cancelled those errors in the moments as given by Sheppard's corrections and in the raw first moment, which are normally corrected by the abruptness corrections. To prove this, one may omit the first frequency and examine the moments of the remainder about its start. The results are as follows:

Moments of the remainder of the frequency about its start.

	True value	Raw moment	With Sheppard	With abruptness correction
$(N-n_1)\,\mu_1'''$	2 050 726	2 063 961	2 063 961	2 049 890
$(N-n_1)\,\mu_2'''$	6 459 788	6 538 164	6 457 077	6 459 401
$(N-n_1)\,\mu_3'''$	25 888 986	26 400 593	25 884 603	25 889 420
$(N-n_1)\,\mu_4'''$	122 960 760	126 209 279	122 968 578	122 966 457

Thus we see that when the first frequency is excluded from consideration, the raw moments are improved by Sheppard's corrections, but the abruptness corrections make a still further improvement in all four moments. Hence the previous unusual result was due entirely to the effect of the first frequency.

But obviously Sheppard's corrections alone cannot always be relied on to give good results where their use is not justifiable, and since using the abruptness corrections for full first range may actually be worse than the raw moments, and give errors of more than 7%, the need for our present method is clearly demonstrated.

Lastly we may inquire whether the value ·3 taken for λ can be improved. Using the four moments obtained by the present method and fitting a Type I curve in the usual way, we obtain the following equation, reduced to the original scale:

$$y = 154{\cdot}956 \times 10^6 \left(\frac{x}{{\cdot}9995}\right)^{{\cdot}9975} \left(1 - \frac{x}{{\cdot}9995}\right)^{10{\cdot}9885}.$$

Distance from start of curve to mean $= 2{\cdot}3793 \times {\cdot}06$, distance from start of second frequency to mean $= 2{\cdot}0477 \times {\cdot}06$, hence $\lambda = {\cdot}3316$ and the base of the first frequency is ·0199.

Hence in fitting a curve it is best not to fix the start at the value given by our tables, but to fit with four moments.

Example 2. The curve in the last example was one in which our tables could reasonably be expected to give a good result, as its equation could be expanded in the form of the auxiliary curve. We will now take a curve whose equation cannot be so expanded,

$$y = y_0 x^{0{\cdot}1} (1 - x)^5.$$

If $N = 10^6$ as before, $y_0 = 7{,}611{,}605{\cdot}925$. Take $h = {\cdot}08$, $\lambda = {\cdot}5$, then $\lambda h = {\cdot}04$.

The "frequencies" are as follows:

x	n_x
0— ·04	180,698
·04— ·12	311,260
·12— ·20	212,991
·20— ·28	134,844
·28— ·36	79,796
·36— ·44	43,790
·44— ·52	21,913
·52— ·60	9,726
·60— ·68	3,663
·68— ·76	1,082
·76— ·84	216
·84— ·92	21
·92—1·0	0
Total	1,000,000

(i) The n_6 test. Table II gives:

when $\lambda = \cdot 4$, $n_6 = 198{,}469$; when $\lambda = \cdot 5$, $n_6 = -19{,}300$.

Hence, since n_6 is really 43,790, $\lambda = \cdot 5$ is the nearest value and this is the correct value.

(ii) The c_1 test. From Table III we have:

when $\lambda = \cdot 8$, $c_1 = -66{,}710$; when $\lambda = \cdot 9$, $c_1 = +6{,}469$.

Hence this test gives $\lambda = \cdot 9$, a bad result.

(iii) The e test. The results are as follows:

Using 3 frequencies, $\lambda = \cdot 24$,
„ 4 „ $\lambda = \cdot 47$,
„ 5 „ $\lambda = \cdot 52$,
„ 6 „ $\lambda = \cdot 50$,
„ 7 „ $\lambda = \cdot 50$.

Hence again four frequencies are sufficient to give the value of λ correct to one figure.

Taking $\lambda = \cdot 5$ and carrying out the work as before, with $h = 1$ as working unit, we get

$$(N - n_1)\,\nu_1''' = 1{,}511{,}250, \qquad (N - n_1)\,\nu_3''' = 19{,}012{,}399{\cdot}5,$$
$$(N - n_1)\,\nu_2''' = 4{,}643{,}887{\cdot}5, \qquad (N - n_1)\,\nu_4''' = 92{,}962{,}029.$$
$$K_1 = -74{,}940{\cdot}3, \qquad K_3 = +3318{\cdot}0,$$
$$K_2 = +14{,}650{\cdot}5, \qquad K_4 = +2832{\cdot}7.$$

$$N\mu_1' = 1{,}511{,}250 - 74{,}940 = 1{,}436{,}310,$$
$$N\mu_2' = 4{,}643{,}887{\cdot}5 - 68{,}275{\cdot}2 + 14{,}650{\cdot}5 = 4{,}590{,}263,$$
$$N\mu_3' = 19{,}012{,}399{\cdot}5 - 377{,}812{\cdot}5 + 3318{\cdot}0 = 18{,}637{,}905,$$
$$N\mu_4' = 92{,}962{,}029 - 2{,}321{,}944 + 23{,}896 + 2833 = 90{,}666{,}814.$$

$$\mu_1' = 1{\cdot}436\,310, \qquad \mu_3' = 18{\cdot}637\,905,$$
$$\mu_2' = 4{\cdot}590\,263, \qquad \mu_4' = 90{\cdot}666\,814.$$

The last three moments are now referred to the mean and give the following comparison with the true values:

	True value	By present method	Error %
Mean from start of second frequency	1·436 620	1·436 310	− ·022
μ_2	2·525 587	2·527 277	+ ·067
μ_3	4·788 490	4·784 959	− ·074
μ_4	27·616 548	27·637 739	+ ·077

Let us again compare these results with those obtained when the first frequency is placed on the same base as the other frequencies, taking the origin at the supposed start of the first frequency. We get

$$\nu_1' = 2{\cdot}420\,901, \qquad \nu_3' = 38{\cdot}319\,701,$$
$$\nu_2' = 8{\cdot}530\,864, \qquad \nu_4' = 203{\cdot}750\,548.$$

These if taken without correction give:

Mean from start of second frequency $= 1{\cdot}420\,901$, error $= -1{\cdot}094\%$,

$\nu_2 = 2{\cdot}670\,102$, error $= +5{\cdot}722\%$,

$\nu_3 = 4{\cdot}739\,218$, error $= -1{\cdot}029\%$,

$\nu_4 = 29{\cdot}616\,321$, error $= +7{\cdot}241\%$.

Applying Sheppard's corrections,

$$\nu_1' = 2{\cdot}420\,901,$$
$$\nu_2' - \tfrac{1}{12} = 8{\cdot}447\,531,$$
$$\nu_3' - \tfrac{1}{4}\nu_1' = 37{\cdot}714\,476,$$
$$\nu_4' - \tfrac{1}{2}\nu_2' + \tfrac{7}{240} = 199{\cdot}514\,283.$$

Taking these as true moment-coefficients and transferring to the mean, we get

$\mu_2 = 2{\cdot}586\,769$, error $= +2{\cdot}422\%$,

$\mu_3 = \nu_3 = 4{\cdot}739\,218$, error $= -1{\cdot}029\%$,

$\mu_4 = 28{\cdot}310\,447$, error $= +2{\cdot}513\%$.

Hence Sheppard's corrections make a considerable improvement in the second and fourth moments, but the errors are still large; the good results which they gave in the first example were due to a good luck which is not repeated here.

The Pairman-Pearson abruptness corrections are

$$H_1 = -\cdot 004\,598, \qquad H_3 = +\cdot 000\,216,$$
$$H_2 = -\cdot 006\,360, \qquad H_4 = +\cdot 005\,776,$$

whence the fully corrected moment-coefficients are

$$\mu_1' = 2{\cdot}425\,499, \qquad \mu_3' = 37{\cdot}714\,692,$$
$$\mu_2' = 8{\cdot}453\,891, \qquad \mu_4' = 199{\cdot}520\,059.$$

The mean is thus 1·425 499 from the start of the second frequency, the error being $-\cdot 774\%$. We get also

$$\mu_2 = 2{\cdot}570\,846, \quad \text{error} = +1{\cdot}792\%,$$
$$\mu_3 = 4{\cdot}738\,621, \quad \text{error} = -1{\cdot}041\%,$$
$$\mu_4 = 28{\cdot}189\,347, \quad \text{error} = +2{\cdot}074\%.$$

These results are an improvement on the results with Sheppard's correction only, but are considerably worse than those obtained by the method of this paper.

Example 3. We will now try our corrections on an asymptotic frequency distribution.

For this purpose let us take the Type III curve

$$y = \frac{10^6}{\Gamma(\cdot 8)} x^{-\cdot 2} e^{-x}.$$

Taking $h = \sqrt{\cdot 8}$, $\lambda = \cdot 5$, we get, from the *Tables of the Incomplete* Γ-*Function*:

x	u	$I(u, p)$	Frequencies
$\frac{1}{2}\sqrt{\cdot 8}$	·5	·466 3930	466,393
$1{\cdot}5\sqrt{\cdot 8}$	1·5	·807 4006	341,008
$2{\cdot}5\sqrt{\cdot 8}$	2·5	·926 7582	119,358
$3{\cdot}5\sqrt{\cdot 8}$	3·5	·971 5767	44,818
$4{\cdot}5\sqrt{\cdot 8}$	4·5	·988 8457	17,269
$5{\cdot}5\sqrt{\cdot 8}$	5·5	·995 5913	6,746
$6{\cdot}5\sqrt{\cdot 8}$	6·5	·998 2487	2,657
$7{\cdot}5\sqrt{\cdot 8}$	7·5	·999 3018	1,053
$8{\cdot}5\sqrt{\cdot 8}$	8·5	·999 7208	419
$9{\cdot}5\sqrt{\cdot 8}$	9·5	·999 8881	167
$10{\cdot}5\sqrt{\cdot 8}$	10·5	·999 9551	67
$11{\cdot}5\sqrt{\cdot 8}$	11·5	·999 9820	27
$12{\cdot}5\sqrt{\cdot 8}$	12·5	·999 9927	11
$13{\cdot}5\sqrt{\cdot 8}$	13·5	·999 9971	4
$14{\cdot}5\sqrt{\cdot 8}$	14·5	·999 9988	2
$15{\cdot}5\sqrt{\cdot 8}$	15·5	·999 9995	1
		Total	1,000,000

(i) The n_6 test. Table II gives:

when $\lambda = \cdot 6$, $n_6 = 101{,}244$; when $\lambda = \cdot 7$, $n_6 = -133{,}346$.

Hence, since n_6 is really 6,746, this test gives $\lambda = \cdot 6$.

(ii) The e test. This test gives:

Using 3 frequencies, $\lambda = \cdot 1$,
" 4 " $\lambda = \cdot 3$,
" 5 " $\lambda = \cdot 4$,
" 6 " $\lambda = \cdot 5$,
" 7 " $\lambda = \cdot 5$.

Hence six frequencies give the correct value of λ.

We find, with $h = 1$ as working unit,

$(N - n_1)\nu_1''' = 579{,}542{\cdot}5$, $(N - n_1)\nu_3''' = 3{,}632{,}652{\cdot}625$,
$(N - n_1)\nu_2''' = 1{,}154{,}247{\cdot}75$, $(N - n_1)\nu_4''' = 15{,}517{,}049{\cdot}4375$.

With $\lambda = \cdot 5$, Table VI gives

$K_1 = -186{,}671{\cdot}9$, $K_3 = -4{,}253{\cdot}5$,
$K_2 = +38{,}430{\cdot}9$, $K_4 = +13{,}794{\cdot}1$.

Applying the full corrections we get

$\mu_1' = \cdot 397\ 8706$, $\mu_3' = 3{\cdot}483\ 5135$,
$\mu_2' = 1{\cdot}148\ 2114$, $\mu_4' = 14{\cdot}969\ 2831$,

whence the moment-coefficients about the mean are

$\mu_2 = \cdot 989\ 910$,
$\mu_3 = 2{\cdot}238\ 961$,
$\mu_4 = 10{\cdot}440\ 733$.

Now h has been taken as 1, whereas it should have been $\sqrt{\cdot 8}$. Hence we multiply μ_s by $(\sqrt{\cdot 8})^s$ and get the following results:

$\mu_1' = \cdot 355\ 8663$, $\mu_3 = 1{\cdot}602\ 070$,
$\mu_2 = \cdot 791\ 928$, $\mu_4 = 6{\cdot}682\ 069$.

Now the true mean is $\cdot 8$ from the start of the curve, or $\cdot 8 - \lambda h = \cdot 352\ 7864$ from the start of the second frequency. Thus the error in the mean as here obtained is $+0{\cdot}87\,\%$.

The true values of μ_2, μ_3 and μ_4 are $\cdot 8$, $1{\cdot}6$ and $6{\cdot}72$ respectively. Hence the errors in our values are $-1{\cdot}01\,\%$, $+0{\cdot}13\,\%$ and $-0{\cdot}56\,\%$ respectively.

These results, while not as good as those obtained for non-asymptotic frequencies, are still very reasonable, especially when it is remembered that nearly half the total frequency is contained in the first group.

We will again compare these results with those obtained by assuming the first sub-range to be equal to the others.

On trying Miss Pearse's method we find no value of q, that is, the method fails to indicate an asymptotic distribution.

The raw moment-coefficients about the supposed start of the first frequency are

$$\nu_1' = 1{\cdot}346\,346, \qquad \nu_3' = 9{\cdot}425\,9295,$$
$$\nu_2' = 2{\cdot}963\,538, \qquad \nu_4' = 39{\cdot}854\,073.$$

The uncorrected moments about the mean work out to be

$$\nu_2 = 1{\cdot}150\,890, \qquad \nu_3 = 2{\cdot}336\,990, \qquad \nu_4 = 11{\cdot}465\,859.$$

Reducing to the base $h = \sqrt{\cdot 8}$, we get:

Mean from start of second frequency $= {\cdot}346\,346 \times \sqrt{\cdot 8} = {\cdot}309\,781$; error $= -12{\cdot}2\,\%$,

$\nu_2 = 1{\cdot}150\,890 \times {\cdot}8 = {\cdot}920\,7120$; error $= +15{\cdot}1\,\%$,
$\nu_3 = 2{\cdot}336\,990 \times (\sqrt{\cdot 8})^3 = 1{\cdot}672\,214$; error $= +4{\cdot}51\,\%$,
$\nu_4 = 11{\cdot}465\,859 \times {\cdot}64 = 7{\cdot}338\,150$; error $= +9{\cdot}20\,\%$.

The percentage errors given are those which would be made if these results were taken instead of the true moment-coefficients.

Now applying Sheppard's corrections, we have

$$\mu_1' = 1{\cdot}346\,346,$$
$$\mu_2' = 2{\cdot}880\,205,$$
$$\mu_3' = 9{\cdot}089\,343,$$
$$\mu_4' = 38{\cdot}375\,221,$$

giving, with base $\sqrt{\cdot 8}$,

$\mu_2 = 1{\cdot}067\,557 \times {\cdot}8 = {\cdot}854\,0456$; error $= +6{\cdot}76\,\%$,
$\mu_3 = 2{\cdot}336\,988 \times (\sqrt{\cdot 8})^3 = 1{\cdot}672\,214$; error $= +4{\cdot}51\,\%$,
$\mu_4 = 10{\cdot}893\,331 \times {\cdot}64 = 6{\cdot}971\,732$; error $= +3{\cdot}75\,\%$.

The mean is as above, with an error of $-12{\cdot}2\,\%$.

Hence Sheppard's corrections make a good improvement in the second and fourth moments.

Applying now the abruptness corrections on the assumption that $h_0 = h$, we obtain

$$H_1 = +{\cdot}032\,0855, \qquad H_3 = +{\cdot}011\,2586,$$
$$H_2 = -{\cdot}003\,6942, \qquad H_4 = -{\cdot}003\,1652.$$

Hence $\mu_1' = \nu_1' - H_1 = 1{\cdot}314\,2605$, and H_1 has again acted the wrong way. The mean is now ${\cdot}314\,2605 \times \sqrt{\cdot 8} = {\cdot}281\,0832$ from the start of the second frequency and is in error to the extent of $20{\cdot}3\,\%$. We have also

$$\mu_2' = 2{\cdot}883\,899, \qquad \mu_3' = 9{\cdot}100\,602, \qquad \mu_4' = 38{\cdot}372\,055,$$

whence, reduced to base $\sqrt{\cdot 8}$, the moment-coefficients about the mean are

$\mu_2 = 1{\cdot}156\,618 \times {\cdot}8 = {\cdot}925\,294$; error $= +15{\cdot}7\,\%$,
$\mu_3 = 2{\cdot}270\,212 \times (\sqrt{\cdot 8})^3 = 1{\cdot}624\,431$; error $= +1{\cdot}53\,\%$,
$\mu_4 = 11{\cdot}467\,132 \times {\cdot}64 = 7{\cdot}338\,964$; error $= +9{\cdot}21\,\%$.

Thus three of these moments are actually worse than the raw moments. The results of this example are shown in the following table.

Moments of Asymptotic Curve $y = \frac{10^6}{\Gamma(\cdot 8)} x^{-\cdot 2} e^{-x}$.

	Mean, from $x = \frac{1}{2}\sqrt{\cdot 8}$	μ_2	μ_3	μ_4
True value ...	·352 786	·8	1·6	6·72
Present method	·355 866	·791 928	1·602 070	6·682 069
Raw moments ...	·309 781	·920 712	1·672 214	7·338 150
With Sheppard ...	·309 781	·854 046	1·672 214	6·971 732
Assumption, $h = h_0$	·281 083	·925 294	1·624 431	7·338 964

Example 4. Age incidence of Whooping Cough cases in Metropolitan Boroughs in 1912:

Under 1 year	212	
1—2 years	427	
2—3 „	344	1463.
3—4 „	270	
4—5 „	210	
5—10 „	253	
10—15 „	14	
Total	1730	

One case between 40 and 45 years has been omitted.

Now here we have the difficulty, common in such statistics, that after the first five years the frequencies are given only for 5-yearly groupings. To get λ from the n_6 test a knowledge of the frequency in the sixth year is required; further, in finding the raw moments omitting the first frequency, the sub-ranges must be equal, and this is impossible as the frequencies stand at present. We have therefore to make an estimate of the frequency in the sixth year, and either arrange the remaining frequency in groups 6—11 years, 11—16 years, etc., or split it entirely into yearly groupings. In the present case, as there is not a long tail, the second method is feasible, and may be performed as follows:

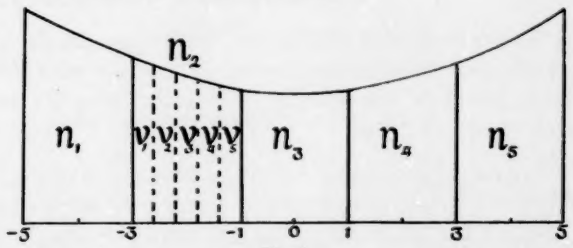


Fig. 6.

Let us take five frequencies n_1, n_2, n_3, n_4, n_5 on equal ranges of 2 units, and fit the parabola

$$\int_0^x y\,dx = ax + bx^2 + cx^3 + dx^4 + ex^5$$

to them. We get, taking the origin as in the figure,

$$\begin{aligned}
n_1 &= \int_{-5}^{-3} y\,dx = 2a - 16b + 98c - 544d + 2882e,\\
n_2 &= 2a - 8b + 26c - 80d + 242e,\\
n_3 &= 2a + 2c + 2e,\\
n_4 &= 2a + 8b + 26c + 80d + 242e,\\
n_5 &= 2a + 16b + 98c + 544d + 2882e,
\end{aligned}$$

whence

$$\begin{aligned}
3840a &= 9n_1 - 116n_2 + 2134n_3 - 116n_4 + 9n_5,\\
384b &= 5n_1 - 34n_2 + 34n_4 - 5n_5,\\
384c &= -n_1 + 12n_2 - 22n_3 + 12n_4 - n_5,\\
768d &= -n_1 + 2n_2 - 2n_4 + n_5,\\
3840e &= n_1 - 4n_2 + 6n_3 - 4n_4 + n_5.
\end{aligned}$$

We now divide n_2 into five frequencies $\nu_1, \nu_2, \nu_3, \nu_4, \nu_5$ on equal bases, getting

$$\begin{aligned}
\nu_1 &= \int_{-3}^{-13/5} y\,dx = \frac{1}{15625}(\ 399n_1 + 4054n_2 - 1931n_3 + 729n_4 - 126n_5),\\
\nu_2 &= \int_{-13/5}^{-11/5} y\,dx = \frac{1}{15625}(\ 69n_1 + 3849n_2 - 1086n_3 + 349n_4 - 56n_5),\\
\nu_3 &= \int_{-11/5}^{-9/5} y\,dx = \frac{1}{15625}(-111n_1 + 3319n_2 - 41n_3 - 56n_4 + 14n_5),\\
\nu_4 &= \int_{-9/5}^{-7/5} y\,dx = \frac{1}{15625}(-181n_1 + 2599n_2 + 1039n_3 - 401n_4 + 69n_5),\\
\nu_5 &= \int_{-7/5}^{-1} y\,dx = \frac{1}{15625}(-176n_1 + 1804n_2 + 2019n_3 - 621n_4 + 99n_5).
\end{aligned}$$

To apply these formulae to the Whooping Cough statistics, we suppose for this purpose that the frequency starts at 0 years, and we have the five frequencies 1463, 253, 14, 0, 0 on equal bases. Then the frequency 253 divides into 101, 68, 43, 26, 15.

If we take the five frequencies 253, 14, 0, 0, 0 and apply the above formulae to split up the 14 frequency, unsatisfactory results are obtained, some of the resulting frequencies being negative. We accordingly take the same five frequencies as before and divide n_3 into $\nu_1', \nu_2', \nu_3', \nu_4', \nu_5'$ by the formulae

$$\begin{aligned}
\nu_1' &= \int_{-1}^{-3/5} y\,dx = \frac{1}{15625}(-126n_1 + 1029n_2 + 2794n_3 - 671n_4 + 99n_5),\\
\nu_2' &= \int_{-3/5}^{-1/5} y\,dx = \frac{1}{15625}(-56n_1 + 349n_2 + 3289n_3 - 526n_4 + 69n_5),
\end{aligned}$$

$$\nu_3' = \int_{-1/5}^{+1/5} y\,dx = \frac{1}{15625}\,(\quad 14n_1 - 181n_2 + 3459n_3 - 181n_4 + 14n_5),$$

$$\nu_4' = \int_{1/5}^{3/5} y\,dx = \frac{1}{15625}\,(\quad 69n_1 - 526n_2 + 3289n_3 + 349n_4 - 56n_5),$$

$$\nu_5' = \int_{3/5}^{1} y\,dx = \frac{1}{15625}\,(\quad 99n_1 - 671n_2 + 2794n_3 + 1029n_4 - 126n_5).$$

These divide the frequency of 14 roughly into 7, 3, 2, 1, 1.

We thus have the revised data:

Under 1 year	212
1—2 years	427
2—3 „	344
3—4 „	270
4—5 „	210
5—6 „	101
6—7 „	68
7—8 „	43
8—9 „	26
9—10 „	15
10—11 „	7
11—12 „	3
12—13 „	2
13—14 „	1
14—15 „	1
Total	1730

(i) The n_6 test. Table II gives:

when $\lambda = \cdot4$, $n_6 = 269{\cdot}840$; when $\lambda = \cdot5$, $n_6 = 30{\cdot}555$.

Hence $\cdot5$ is the nearest value for λ.

(ii) The c_1 test. Table III gives:

when $\lambda = \cdot8$, $c_1 = -33{\cdot}903$; when $\lambda = \cdot9$, $c_1 = +53{\cdot}638$.

Hence this gives $\lambda = \cdot8$.

(iii) The e test gives the following results:

Using 3 frequencies, $\lambda = \cdot3$,
„ 4 „ $\lambda = \cdot4$,
„ 5 „ $\lambda = \cdot5$,
„ 6 „ $\lambda > 1{\cdot}0$,
„ 7 „ $\lambda > 1{\cdot}5$.

We have to choose from these the best value of λ, i.e. the value which gives the best fit to the total frequency. We may reject, perhaps, in this case the values of λ greater than unity, but there is no guide as to which of the other values is best.

Accordingly, curves have been fitted for the values ·5 and ·8, using four moments. In each case a Type I curve is indicated by the β's and the value ·8 gives a better fit than the value ·5.

The fitting for $\lambda = \cdot 8$ is, with the usual notation, as follows:

About the start of the second frequency,

$$(N-n_1)\nu_1''' = 3717, \qquad (N-n_1)\nu_3''' = 83873{\cdot}25,$$
$$(N-n_1)\nu_2''' = 15171{\cdot}5, \qquad (N-n_1)\nu_4''' = 570882{\cdot}875.$$
$$K_1 = -99{\cdot}9286, \qquad K_3 = -3{\cdot}7007,$$
$$K_2 = +30{\cdot}0474, \qquad K_4 = +7{\cdot}5620,$$

whence, using the full corrections,

$$\mu_1' = 2{\cdot}090\,793, \qquad \mu_2' = 8{\cdot}713\,900,$$
$$\mu_3' = 47{\cdot}942\,369, \qquad \mu_4' = 325{\cdot}635\,238.$$

These give

$$\mu_2 = 4{\cdot}342\,485,$$
$$\mu_3 = 11{\cdot}564\,934,$$
$$\mu_4 = 95{\cdot}909\,587.$$
$$\beta_1 = 1{\cdot}633\,320, \qquad \beta_2 = 5{\cdot}086\,106.$$

These values indicate a Type I curve. Fitting with four moments, we have

$$r = \frac{6\beta_2 - 6\beta_1 - 6}{3\beta_1 - 2\beta_2 + 6} = 20{\cdot}22227,$$

$$\epsilon = \frac{r^2}{4 + \dfrac{(r+2)^2}{4(r+1)}\beta_1} = 30{\cdot}28830,$$

$$b = r\sqrt{\frac{\mu_2(r+1)}{\epsilon}} = 35{\cdot}27423.$$

m_1+1, m_2+1 are the roots of the quadratic $m'^2 - rm' + \epsilon = 0$, whence

$$m_1 = 0{\cdot}62899, \qquad m_2 = 17{\cdot}59328.$$

$$a_1 = \frac{bm_1}{m_1+m_2} = 1{\cdot}217584, \qquad a_2 = \frac{bm_2}{m_1+m_2} = 34{\cdot}05665^{-},$$

$$y_0 = \frac{N}{b}\,\frac{m_1^{m_1} m_2^{m_2}}{(m_1+m_2)^{m_1+m_2}}\,\frac{\Gamma(m_1+m_2+2)}{\Gamma(m_1+1)\,\Gamma(m_2+1)} = 425{\cdot}880.$$

$$\text{Mean to mode} = \frac{1}{2}\,\frac{r+2}{r-2}\,\frac{\mu_3}{\mu_2} = 1{\cdot}623\,905.$$

The equation, referred to the mode, is

$$y = 425{\cdot}880\left(1 + \frac{x}{1{\cdot}217\,584}\right)^{0{\cdot}62899}\left(1 - \frac{x}{34{\cdot}05665}\right)^{17{\cdot}59328}.$$

This curve is drawn in the diagram (Fig. 7).

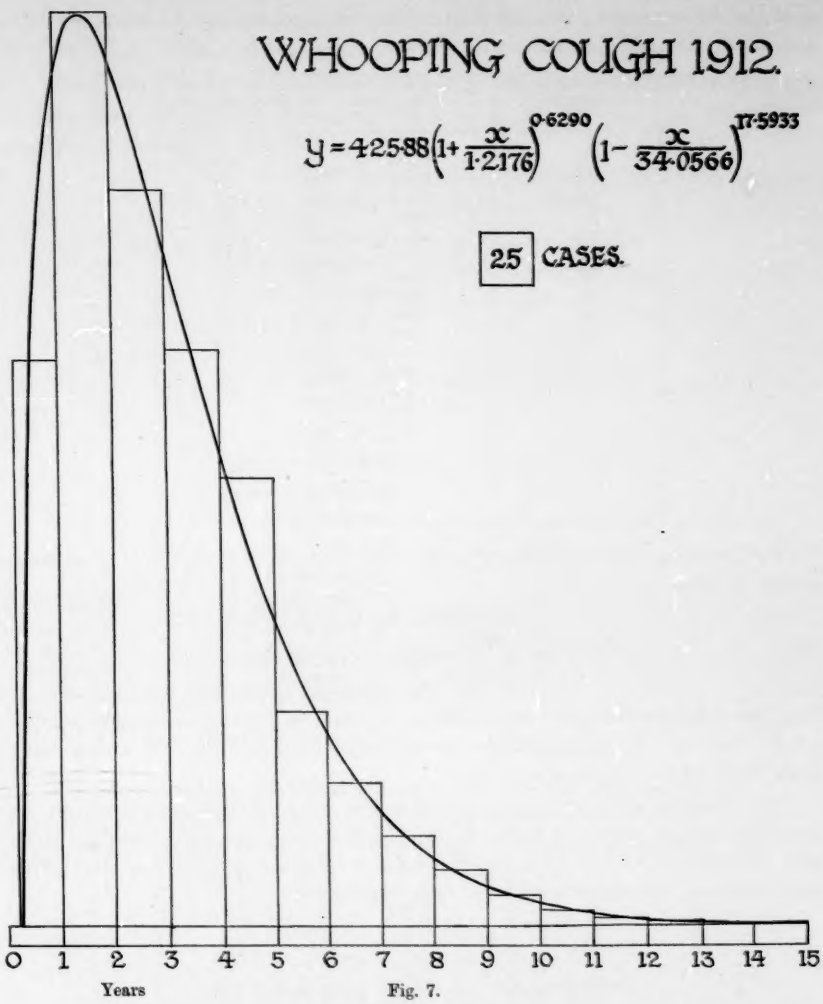

WHOOPING COUGH 1912.
$y = 42588\left(1+\frac{x}{1{\cdot}2176}\right)^{0{\cdot}6290}\left(1-\frac{x}{34{\cdot}0566}\right)^{17{\cdot}5933}$
25 CASES.
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Years

Fig. 7.

To find the theoretical frequencies, the mid-ordinates of the groups (except the first) were calculated, and the frequencies were calculated from the formula

$$n_r = \frac{h}{5760}\{5178y_r + 308\,(y_{r-1} + y_{r+1}) - 17\,(y_{r-2} + y_{r+2})\}.$$

The first two frequencies were calculated from the formula

$$\int_0^{x'} y\,dx' = y_0\left(\frac{b}{a_2}\right)^{m_2}\left(\frac{x'}{a_1}\right)^{m_1+1} a_1\left[\frac{1}{m_1+1} - \frac{m_2}{m_1+2}\left(\frac{x'}{b}\right) + \frac{m_2\,(m_2-1)}{2\,!\,(m_1+3)}\left(\frac{x'}{b}\right)^2 - \ldots\right]$$

obtained by expanding the equation of the curve referred to its start as origin, and integrating. The results are as follows:

n observed	n theoretical	Contribution to χ^2
212	213·6	·01
427	418·2	·19
344	363·5	1·05
270	268·0	·01
210	182·1	4·27
101	117·1	2·21
68	72·1	·23
43	42·8	·00
26	24·5	·09
15	13·6	·14
7	7·3	
3	3·9	
2 } 14	1·9 } 14·4	·01
1	0·9	
1	0·4	
		$\chi^2 = 8·21$

For goodness of fit, this gives $P = ·61$. The goodness of fit to the original seven groups is given by $P = ·36$.

Now $$d = \text{mean} - \text{mode} = 1·623\,905,$$

and $$AO \text{ (Fig. 1)} = \text{base of 1st frequency}$$
$$= a_1 - \mu_1' + d = ·750\,696.$$

Now this indicates that the value ·7507 for λ would be a better one to take; if this is done, and the K's evaluated by linear interpolation in Table VI, and a curve again fitted by four moments, a new value of λ is found.

The question naturally arises whether the series of values of λ found by successive fittings tend to a limit; further, is this limit eventually attained if we start off with a bad value of λ, e.g. the value ·5 obtained from the n_6 test? This point has been investigated, with the following results:

$c_1 = 0$ gives $\lambda_1 = ·8$ (nearest).

1st fitting with $\lambda_1 = ·8$ gives $\lambda_2 = ·7507$,
2nd fitting with $\lambda_2 = ·7507$ gives $\lambda_3 = ·7421$,
3rd fitting with $\lambda_3 = ·7421$ gives $\lambda_4 = ·7406$,
4th fitting with $\lambda_4 = ·7406$ gives $\lambda_5 = ·7403$,
5th fitting with $\lambda_5 = ·7403$ gives $\lambda_6 = ·7403$.

The n_6 test gives $\lambda_1 = ·5$ (nearest).

1st fitting with $\lambda_1 = ·5$ gives $\lambda_2 = ·6586$,
2nd fitting with $\lambda_2 = ·6586$ gives $\lambda_3 = ·7198$,
3rd fitting with $\lambda_3 = ·7198$ gives $\lambda_4 = ·7367$,
4th fitting with $\lambda_4 = ·7367$ gives $\lambda_5 = ·7396$,
5th fitting with $\lambda_5 = ·7396$ gives $\lambda_6 = ·7402$.

Hence we may conclude that the values of λ obtained by successive fittings tend to a limit, here ·7403, and this is the case even if a bad value is originally selected.

These data therefore indicate that Whooping Cough starts at ·2597 year of age.

The curve fitted with the value ·7403 for λ should thus give the best fit possible; the details of this fitting are:

$K_1 = -98{\cdot}7699$, $K_2 = +28{\cdot}5936$,
$K_3 = -2{\cdot}4606$, $K_4 = +6{\cdot}6076$.
$\mu_1' = 2{\cdot}091\,462$, $\mu_2' = 8{\cdot}713\,060$,
$\mu_3' = 47{\cdot}943\,086$, $\mu_4' = 325{\cdot}634\,686$,
$\mu_2 = 4{\cdot}338\,847$,
$\mu_3 = 11{\cdot}570\,984$,
$\mu_4 = 95{\cdot}825\,542$.
$\beta_1 = 1{\cdot}639\,145$, $\beta_2 = 5{\cdot}090\,175$,
$r = 19{\cdot}95181$,
$\epsilon = 29{\cdot}65197$,
$m_1 = {\cdot}617\,275$,
$m_2 = 17{\cdot}334\,535$,
$b = 34{\cdot}93440$,
$a_1 = 1{\cdot}201\,223$,
$a_2 = 33{\cdot}73318$,
$y_0 = 427{\cdot}027$,
$d = 1{\cdot}630\,527$,
AO (Fig. 1) = ·740 288.

The equation, referred to the mode, is

$$y = 427{\cdot}027\left(1+\frac{x}{1{\cdot}201\,223}\right)^{0{\cdot}617275}\left(1-\frac{x}{33{\cdot}73318}\right)^{17{\cdot}3345}.$$

The frequencies are calculated as before, with the following results:

n observed	n theoretical	Contribution to χ^2
212	212·5	·00
427	419·2	·15
344	363·7	1·07
270	267·9	·02
210	182·0	4·31
101	117·1	2·21
68	72·1	·23
43	42·8	·00
26	24·6	·08
15	13·6	·14
7	7·3	
3	3·8	
2 } 14	1·9 } 14·3	·01
1	0·9	
1	0·4	
		$\chi^2 = 8{\cdot}22$, $n' = 11$

This gives $P = \cdot 61$. For the original seven frequencies P is still $\cdot 36$. Hence the result of our successive fittings has been to improve very slightly the fit at the start of the curve, but without any sensible increase of P. The conclusion we may draw is that if the start of the frequency obtained by fitting with four moments does not differ greatly from the value selected from the tables without interpolation, it is not worth while to make a second fitting. If this is not the case, then successive fittings steadily improve the goodness of fit; thus the above curve is much better than the one which would be obtained by using the value $\lambda = \cdot 5$ obtained from the n_6 test.

The curve with $\lambda = \cdot 7403$ is shown in Fig. 8.

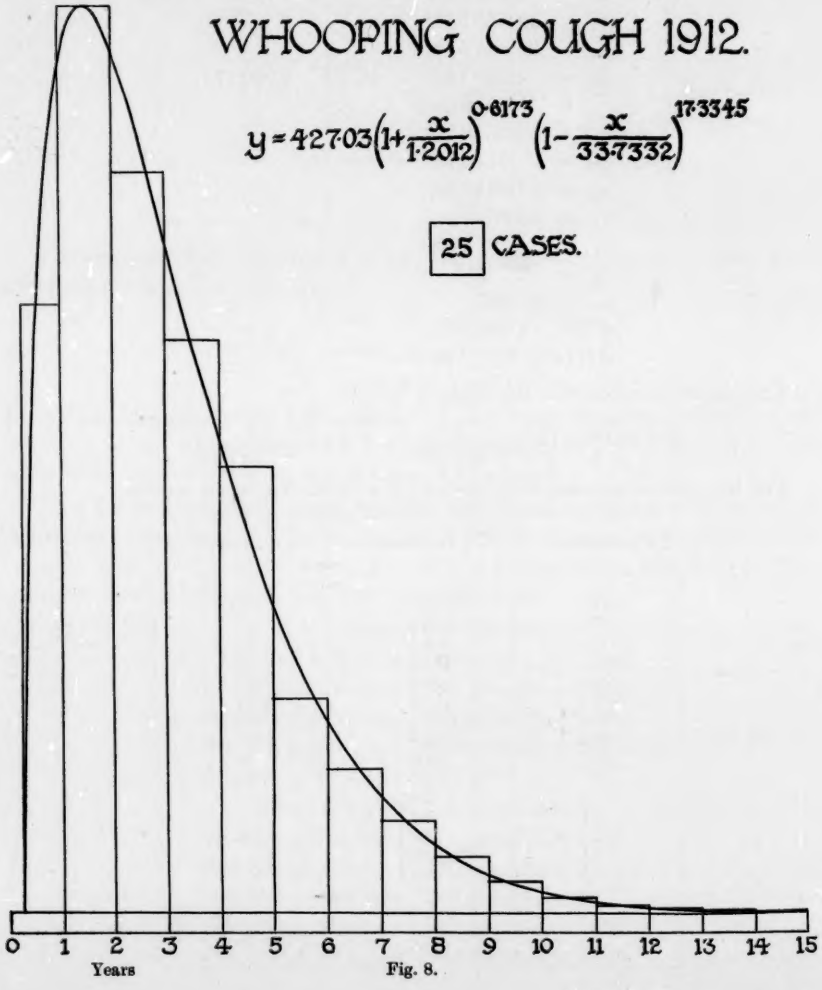

WHOOPING COUGH 1912.
$y = 427{\cdot}03\left(1+\frac{x}{1{\cdot}2012}\right)^{0{\cdot}6173}\left(1-\frac{x}{33{\cdot}7332}\right)^{17{\cdot}3345}$
25 CASES.
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Years

Fig. 8.

Example 5. Dwelling houses in England and Wales first assessed for Income Tax in 1929—30.

Valuation	No. of Houses
Under £20	40,272
20—40	104,243
40—60	27,589
Over 60	12,226

The information here is meagre; no indication is given of the minimum valuation taken, and all houses valued at more than £60 are grouped together. The first point, namely the start of the frequency, must be discovered by our present method. The second difficulty can be met by a hypothetical spreading out of the frequency 12,226 into a suitable "tail."

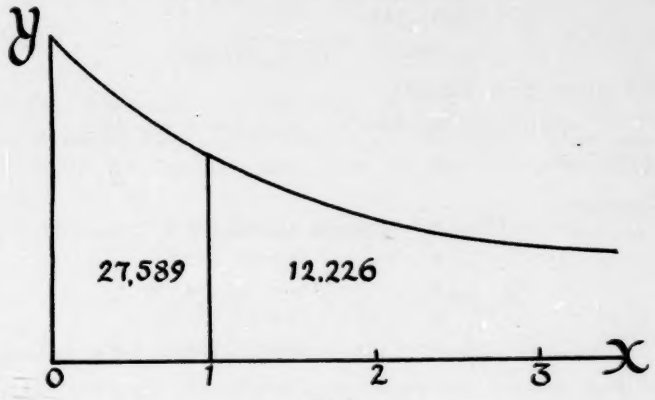


Fig. 9.

Let us assume the form $y = \frac{N'}{\sigma} e^{-x/\sigma}$ for the tail, the origin being as in the figure.

Then $\int_0^x y\,dx = N'(1 - e^{-x/\sigma})$, and we have

$$\int_0^1 y\,dx = 27{,}589 = N'(1 - e^{-1/\sigma}),$$

$$\int_0^\infty y\,dx = 39{,}815 = N',$$

whence $$\frac{1}{\sigma} = 1{\cdot}180\,679,$$

and $$\int_0^x y\,dx = 39{,}815\,(1 - e^{-1{\cdot}180679x}).$$

Putting $x = 2, 3, 4$, etc. we get the corresponding frequencies which are shown in the following distribution:

Valuation	No. of Houses	
Under £20	40,272	
20— 40	104,243	
40— 60	27,589	
60— 80	8,472	12,226
80—100	2,602	
100—120	799	
120—140	245	
140—160	75	
160—180	23	
180—200	7	
200—220	2	
220—240	1	
Total	184,330	

(i) The n_6 test. From Table II,

when $\lambda = \cdot1$, $n_6 = 868{,}753$; when $\lambda = \cdot2$, $n_6 = 247$.

Hence this test gives $\lambda = \cdot2$.

(ii) The e test.

Using 3 frequencies, $\lambda < \cdot05$, $q < \cdot4$,
" 4 " $\lambda = \cdot05$, $q = \cdot5$,
" 5 " $\lambda = \cdot1$, $q = \cdot7$,
" 6 " $\lambda = \cdot1$, $q = \cdot7$,
" 7 " $\lambda = \cdot1$, $q = \cdot7$.

The values of q indicate an asymptotic curve, therefore the $c_1 = 0$ test was not tried.

We find that the value $\lambda = \cdot2$ gives the best results. The e test is remarkably constant for 5, 6 and 7 frequencies, but this is due to the fact that all the frequencies except the first two are already smoothed.

With $\lambda = \cdot2$ we have

$K_1 = -18{,}898{\cdot}41$, $K_3 = +4{,}045{\cdot}02$,
$K_2 = -1{,}092{\cdot}24$, $K_4 = +1{,}550{\cdot}88$.

With the usual notation,

$(N - n_1)\,\nu_1''' = 129{,}484$, $(N - n_1)\,\nu_3''' = 501{,}121$,
$(N - n_1)\,\nu_2''' = 201{,}810{\cdot}5$, $(N - n_1)\,\nu_4''' = 1{,}691{,}053{\cdot}625$.

Applying the full corrections we get the following moment-coefficients about the start of the second frequency:

$\mu_1' = \cdot599\,9327$, $\mu_3' = 2{\cdot}564\,9380$,
$\mu_2' = 1{\cdot}023\,7803$, $\mu_4' = 8{\cdot}657\,8470$,

leading to

$$\mu_2 = \cdot 663\,8611,$$
$$\mu_3 = 1\cdot 154\,1947,$$
$$\mu_4 = 4\cdot 324\,9300,$$
$$\beta_1 = 4\cdot 553\,304, \qquad \beta_2 = 9\cdot 813\,515.$$

Now $2\beta_2 - 3\beta_1 - 6 = -\cdot 032\,882$, which is sufficiently near to zero to allow a Type III curve to be taken.

If the equation referred to the start of the curve is

$$y = y_0 \left(\frac{x}{a}\right)^p e^{-(p+1)\frac{x}{a}},$$

we have

$$p = \frac{4}{\beta_1} - 1 = -\cdot 121\,5170,$$

$$a = 2\sqrt{\frac{\mu_2}{\beta_1}} = \cdot 763\,6693,$$

$$y_0 = \frac{N}{a}\frac{(1+p)^{p+1}}{\Gamma(p+1)} = 198{,}236\cdot 8.$$

Start of curve to mean $= a = \cdot 763\,6693$.

Start of second frequency to mean $= \mu_1' = \cdot 599\,9327$, whence base of first frequency $= \cdot 163\,7366$, and the curve starts at $\cdot 836\,2634 \times £20$, or about £16·7 per annum.

To get the theoretical frequencies, we have

$$\int_0^x y\,dx = N\frac{\Gamma_{(p+1)\frac{x}{a}}(p+1)}{\Gamma(p+1)} = NI(u, p)$$

in the notation of the *Tables of the Incomplete* Γ*-Function*, where

$$u = \frac{(p+1)\frac{x}{a}}{\sqrt{p+1}} = \sqrt{p+1}\,\frac{x}{a} = \cdot 937\,2742\,\frac{x}{a}.$$

The tables mentioned give the following theoretical frequencies which are placed beside the observed frequencies for comparison.

Observed		Theoretical	
40,272		40,858	
104,243		102,972	
27,589		28,352	
8,472		8,441	
2,602		2,566	
799		788	
245		243·5	
75	12,226	75·5	12,147·5
23		23·5	
7		7	
2		2	
1		1	

These results are in fairly good accord but the probability P is zero to five places of decimals. This is explained by the fact that the given frequencies are so large that they would have to be reproduced with extreme fidelity in order to give a reasonable value of P.

It may be remarked that an attempt to fit a curve by four moments, assuming the base of the first frequency to be the same as the others and using the ordinary abruptness corrections, fails owing to the appropriate Type I curve omitting the first frequency.

The curve obtained above is shown in Fig. 10.

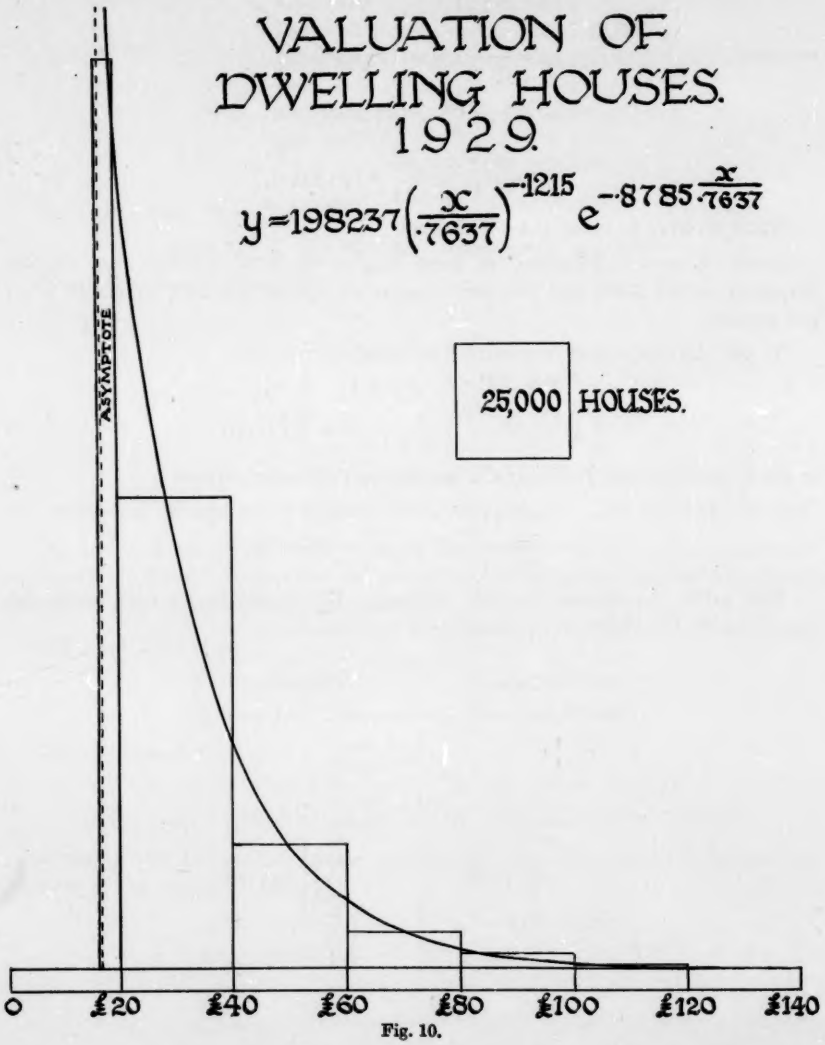


Fig. 10.

Example 6. Cricket Scores, 1931. In order to make the scores as homogeneous as possible, only First Class County Championship matches have been considered and all "not outs" and uncompleted innings ("declared" or abandoned) have been excluded. The distribution is as follows:

Score (Runs)	Number of Innings		Score (Runs)	Number of Innings
0	701		81— 90	35
1, 2	627		91—100	22
3, 4	470		101—110	21
5, 6	374	2109	111—120	15
7, 8	335		121—130	10
9, 10	303		131—140	8
11—20	895		141—150	3
21—30	524		151—160	6
31—40	330		161—170	1
41—50	170		171—180	0
51—60	105		181—190	1
61—70	83		191—200	0
71—80	56		201—210	1
			Total	5096

(i) The n_6 test gives, for $\lambda = \cdot 8$, $n_6 = 419$;

$\lambda = \cdot 9$, $n_6 = 284$.

Hence, since $n_6 = 303$, $\lambda = \cdot 9$ is the nearest value.

(ii) The e test gives the following results:

Using 3 frequencies, $\lambda = \cdot 5$ $(q = \cdot 6)$,

„ 4 „ $\lambda = \cdot 7$ $(q = \cdot 8)$,

„ 5 „ $\lambda > 1$ $(q > 1)$.

Now we can see from the data that λ should not be greater than $\cdot 5$, hence we take the value $\cdot 5$ given by three frequencies in the e test*. The value of q indicates that the distribution is asymptotic.

* This choice is supported by results obtained if we do not use the subdivision of the group 1—10 runs, i.e., if we take as the first seven frequencies 701, 2109, 895, 524, 330, 170, 105. With these the e test gives:

Using 3 frequencies, $\lambda = \cdot 05$;

„ 4 „ $\lambda = \cdot 1$;

„ 5 „ $\lambda = \cdot 1$;

„ 6 „ $\lambda = \cdot 1$;

„ 7 „ $\lambda = \cdot 2$.

Hence, since our sub-ranges are now five times as great, the value $\lambda = \cdot 1$ given by 4, 5 and 6 frequencies corresponds to the value $\lambda = \cdot 5$ taken above. The e test here gives more constant results as the frequencies are more regular than those in the subdivisions.

Table VI cannot be used in this example, as h_0 is not equal to h. We get, from Table IV, taking $h = 1$, $h_0 = \cdot 2$, $\lambda = \cdot 5$,

$$\frac{1}{h_0}n_1\mu_1'' = -195{\cdot}362, \quad \therefore\ n_1\mu_1'' = -39{\cdot}072.$$

$$\frac{1}{h_0^2}n_1\mu_2'' = +\ 68{\cdot}727, \quad \therefore\ n_1\mu_2'' = +\ 2{\cdot}749.$$

$$\frac{1}{h_0^3}n_1\mu_3'' = -\ 26{\cdot}624, \quad \therefore\ n_1\mu_3'' = -\ 0{\cdot}213.$$

$$\frac{1}{h_0^4}n_1\mu_4'' = +\ 10{\cdot}884, \quad \therefore\ n_1\mu_4'' = +\ 0{\cdot}017.$$

Table V gives

$$a_1 = -\ 977{\cdot}525^{+}, \quad \therefore\ a_1' = \ \ 5a_1 = -4{,}887{\cdot}63.$$
$$a_2 = +1{,}266{\cdot}217, \quad \therefore\ a_2' = \ 25a_2 = +31{,}655{\cdot}4.$$
$$a_3 = -2{,}260{\cdot}555, \quad \therefore\ a_3' = 125a_3 = -282{,}569.$$
$$a_4 = +2{,}510{\cdot}921, \quad \therefore\ a_4' = 625a_4 = +1{,}569{,}326.$$
$$a_5 = -1{,}253{\cdot}460, \quad \therefore\ a_5' = 3125a_5 = -3{,}917{,}063.$$

The abruptness functions are

$$\frac{a_1'}{12} - \frac{a_3'}{720} + \frac{a_5'}{30240} = -144{\cdot}378.$$

$$-\frac{a_2'}{120} + \frac{a_4'}{3024} = +255{\cdot}162.$$

$$-\frac{a_1'}{40} + \frac{a_3'}{504} - \frac{a_5'}{9600} = -\ 30{\cdot}435.$$

$$\frac{a_2'}{126} - \frac{a_4'}{1440} = -838{\cdot}577.$$

Now to calculate the moments of the remainder about *its* start, when the first frequency is omitted, we must group together the next five frequencies with a total 2109 in order to have all the ranges equal, as required by the theory. We get, with $h = 1$,

$$(N - n_1)\,\nu_1''' = \ 8{,}488{\cdot}5, \qquad (N - n_1)\,\nu_3''' = \ 282{,}622{\cdot}125,$$
$$(N - n_1)\,\nu_2''' = 38{,}074{\cdot}75, \qquad (N - n_1)\,\nu_4''' = 2{,}801{,}954{\cdot}6875,$$

whence, using Equations II on p. 15,

$$\mu_1' = 1{\cdot}629\ 719, \qquad \mu_3' = \ 55{\cdot}037\ 157,$$
$$\mu_2' = 7{\cdot}450\ 238, \qquad \mu_4' = 545{\cdot}958\ 976.$$

Transferring to the mean, we obtain

$$\mu_2 = \ \ 4{\cdot}794\ 254,$$
$$\mu_3 = \ 27{\cdot}268\ 791,$$
$$\mu_4 = 284{\cdot}742\ 098,$$

whence $\beta_1 = 6{\cdot}747\,900, \qquad \beta_2 = 12{\cdot}388\,240.$

These values indicate a Type I curve. When we endeavour, however, to fit a curve by four moments, we find that the curve just omits the first frequency. We therefore fix the start at the point given by $\lambda = {\cdot}5$ and fit with three moments. We have:

$$c = \text{distance of mean from start} = {\cdot}1 + 1{\cdot}629\,719 = 1{\cdot}729\,719,$$

$$r = \frac{2(\mu_2 c^2 + \mu_3 c - \mu_2^2)}{2\mu_2^2 - \mu_3 c} = -64{\cdot}339\,415.$$

The negative value of r indicates that the type has changed to Type VI as a result of fixing the start. The working for a Type VI curve with fixed start is as follows:

$$\gamma_2 = \frac{\mu_2}{c^2} = 1{\cdot}602\,39628,$$

$$\gamma_3 = \frac{\mu_3}{2\mu_2 c} = 1{\cdot}644\,14188.$$

If the equation is $y = y_0 \dfrac{(x-a)^{q_2}}{x^{q_1}}$,

then $$q_2 - q_1 = \frac{1 - 3\gamma_2 + 4\gamma_3}{\gamma_2 - \gamma_3} = -66{\cdot}339\,415,$$

$$q_2 + q_1 = \frac{\gamma_3(1+\gamma_2)(\gamma_2 - 1 - 2\gamma_3)}{(2\gamma_2 - \gamma_3 + \gamma_2\gamma_3)(\gamma_2 - \gamma_3)} = 65{\cdot}619\,867.$$

$$\therefore\ q_1 = 65{\cdot}979\,641,$$

$$q_2 = -\,{\cdot}359\,774,$$

$$a = -\frac{cr}{q_2 + 1} = 173{\cdot}8278,$$

$$y_0 = \frac{N}{a^{r-1}} \frac{\Gamma(q_1)}{\Gamma(1-r)\,\Gamma(q_2+1)} = 1{\cdot}228\,559 \times 10^{151}.$$

Hence the equation is

$$y = 1{\cdot}228\,559 \times 10^{151} \frac{(x - 173{\cdot}8278)^{-359\,774}}{x^{65{\cdot}979\,641}}.$$

The first seven frequencies can be calculated from the formula

$$\int_a^x y\,dx = y_0\, a^{r-1}\, z^{q_2+1}\left[\frac{1}{q_2+1} + \frac{r}{q_2+2}\,z + \frac{r(r+1)}{2!(q_2+3)}\,z^2 + \ldots\right],$$

where $z = 1 - a/x, \qquad y_0 a^{r-1} = 52{,}640{\cdot}23.$

The remaining frequencies can be calculated from the formula

$$n_r = \frac{h}{5760}\left[5178 y_r + 308\,(y_{r-1} + y_{r+1}) - 17\,(y_{r-2} + y_{r+2})\right].$$

The results are as follows:

Score	Observed Frequency	Theoretical Frequency	χ^2
0	701	682·6	0·496
1, 2	627	657·1	1·379
3, 4	470	466·0	0·034
5, 6	374	372·1	0·010
7, 8	335	310·7	1·901
9, 10	303	265·7	5·236
11— 20	895	913·0	0·355
21— 30	524	522·6	0·004
31— 40	330	319·3	0·359
41— 50	170	201·5	4·924
51— 60	105	129·7	4·704
61— 70	83	84·7	0·034
71— 80	56	55·9	0·000
81— 90	35	37·2	0·130
91—100	22	24·9	0·338
101—110	21	16·8	1·050
111—120	15	11·4	1·137
121—130	10	7·7	
131—140	8	5·3	
141—150	3	3·6	
151—160	6	2·5	
161—170	1	1·7	
171—180	0 } 30	1·2 } 24·8	1·090
181—190	1	0·8	
191—200	0	0·6	
201—210	1	0·4	
over 210	0	1·0	
Totals	5096	5096	23·181

Hence, since $n' = 18$, $P = \cdot 144$ from Palin Elderton's Table.

The curve has been drawn in Figs. 11 (*a*) and 11 (*b*). Owing to the large frequencies on small ranges at the start, the distribution cannot be adequately represented in a single diagram. The first figure therefore gives the first six groups and the second figure gives the remaining groups on a reduced scale. The points A in the two diagrams correspond. It is seen that the observed frequencies are most irregular and therefore the curve fits worst in the neighbourhood of 10 and of 50 runs. This can probably be explained on psychological grounds.

5. *Conclusion.* To the question "Where does a given frequency distribution start?" we can give only a qualified answer. Three tests for λ are suggested in this paper and they may give conflicting results. If, however, the distribution be a theoretical one, i.e., if there are no sampling or experimental errors, then the e test should always be used, as it can be relied on to give the start correctly. In other cases the choice of the best value may be guided from some *a priori* knowledge of the distribution or it may have to be made by fitting curves for each value of λ and selecting that which fits best. In all the examples we have chosen, at least one of the tests has given a suitable value of λ. When the best value of λ has been found, the tables at the end of this paper provide adequate corrections for the

CRICKET SCORES 1931.

$$Y = 1{\cdot}22856 \times 10^{151} \frac{(X - 173{\cdot}828)^{-{\cdot}359774}}{X^{65{\cdot}9796}}$$

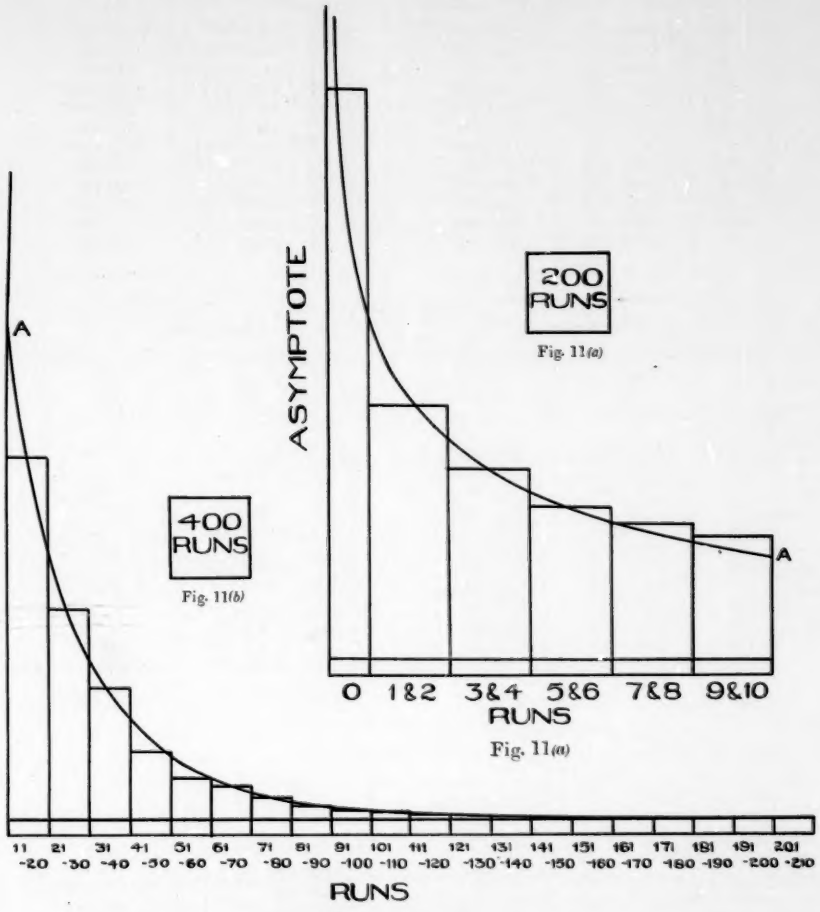


Fig. 11(*a*)

Fig. 11(*b*)

Fig. 11(*b*)

moments. The best results will usually be obtained by fitting with four moments, but if this fails, the start of the curve should be fixed at the point given by the chosen value of λ.

I have to thank Professor Karl Pearson, F.R.S., under whose guidance the work was carried out, for his continual help and stimulating advice.

TABLE I. (*Exponential Test.*)

3 frequencies		
λ	$\frac{{}_3L_1}{{}_3L_2}$	${}_3L_1$
·05	4·550 518	1·322 219
·1	3·708 312	1·041 393
·2	2·956 036	·778 151
·3	2·570 064	·636 822
·4	2·324 251	·544 068
·5	2·150 660	·477 121
·6	2·020 213	·425 969
·7	1·917 981	·385 351
·8	1·835 378	·352 183
·9	1·767 062	·324 511
1·0	1·709 511	·301 030
1·5	1·518 181	·221 849

4 frequencies		
λ	$\frac{{}_4L_1}{{}_4L_2}$	${}_4L_1$
·05	8·741 460	1·031 655$^-$
·1	6·809 970	·760 566
·2	5·122 760	·514 910
·3	4·275 215$^-$	·389 038
·4	3·743 076	·309 985$^-$
·5	3·371 237	·255 273
·6	3·094 107	·215 115$^+$
·7	2·878 355$^+$	·184 436
·8	2·704 980	·160 297
·9	2·562 247	·140 867
1·0	2·442 475$^-$	·124 939
1·5	2·047 415$^+$	·075 721

5 frequencies		
λ	$\frac{{}_5L_1}{{}_5L_2}$	${}_5L_1$
·05	13·312 918	·913 636
·1	10·144 490	·648 882
·2	7·414 998	·414 396
·3	6·061 800	·298 039
·4	5·219 689	·227 169
·5	4·635 132	·179 552
·6	4·201 718	·145 591
·7	3·865 712	·120 359
·8	3·596 643	·101 037
·9	3·375 784	·085 889
1·0	3·190 921	·073 786
1·5	2·584 325$^-$	·038 737

6 frequencies		
λ	$\frac{{}_6L_1}{{}_6L_2}$	${}_6L_1$
·05	18·150 433	·845 008
·1	13·640 942	·584 918
·2	9·794 426	·358 510
·3	7·905 153	·248 872
·4	6·736 800	·183 648
·5	5·929 568	·140 815$^-$
·6	5·333 254	·110 941
·7	4·872 341	·089 224
·8	4·504 169	·072 945$^-$
·9	4·202 601	·060 446
1·0	3·950 642	·050 662
1·5	3·126 988	·023 748

7 frequencies		
λ	$\frac{{}_7L_1}{{}_7L_2}$	${}_7L_1$
·05	23·190 383	·798 453
·1	17·259 963	·542 038
·2	12·239 659	·321 906
·3	9·791 384	·217 390
·4	8·284 568	·156 387
·5	7·247 196	·117 067
·6	6·483 021	·090 139
·7	5·893 711	·070 912
·8	5·423 872	·056 750$^-$
·9	5·039 654	·046 063
1·0	4·719 088	·037 839
1·5	3·674 188	·016 153

TABLE II. n_6.

λ	$+n_1$	$-n_2$	$+n_3$	$-n_4$	$+n_5$
0	∞	11·416 667	13·583 333	11·416 667	5·250 000
·1	40·871 795$^+$	10·128 205$^-$	13·053 614	11·232 101	5·219 512
·2	16·910 173	9·089 827	12·576 840	11·059 524	5·190 476
·3	9·427 714	8·238 952	12·145 663	10·897 815$^+$	5·162 791
·4	5·968 296	7·531 704	11·754 011	10·745 989	5·136 364
·5	4·063 492	6·936 508	11·396 825$^+$	10·603 175$^-$	5·111 111
·6	2·903 196	6·430 137	11·069 863	10·468 599	5·086 957
·7	2·147 687	5·995 170	10·769 535$^+$	10·341 576	5·063 830
·8	1·631 683	5·618 317	10·492 794	10·221 491	5·041 667
·9	1·266 271	5·289 285$^+$	10·237 031	10·107 797	5·020 408
1·0	1·000 000	5·000 000	10·000 000	10·000 000	5·000 000

TABLE III. c_1.

λ	$-n_1$	$+n_2$	$-n_3$	$+n_4$	$-n_5$
0	∞	0	0	0	0
·1	11·951 764 469 467	·281 610 530 533	·122 798 560 376	·043 641 915 814	·007 277 439 024
·2	6·838 474 025 974	·553 525 974 026	·267 807 359 307	·098 738 095 238	·016 761 904 762
·3	5·072 935 153 825	·819 856 512 842	·435 412 717 928	·166 119 890 768	·028 683 139 535
·4	4·152 342 755 284	1·083 657 244 716	·626 057 040 998	·246 609 625 668	·043 272 727 273
·5	3·574 603 174 603	1·347 271 825 397	·840 228 174 603	·341 021 825 397	·060 763 888 889
·6	3·171 451 133 408	1·612 548 866 592	1·078 451 133 408	·450 164 251 208	·081 391 304 348
·7	2·870 213 320 777	1·880 983 107 795	1·341 281 598 087	·574 838 772 283	·105 390 957 447
·8	2·634 189 640 769	2·153 810 359 231	1·629 300 751 880	·715 842 105 263	·133 000 000 000
·9	2·442 746 375 855	2·432 073 068 589	1·943 111 141 937	·873 966 444 270	·164 456 632 653
1·0	2·283 333 333 333	2·716 666 666 667	2·283 333 333 333	1·050 000 000 000	·200 000 000 000

c_2.

λ	$+n_1$	$-n_2$	$+n_3$	$-n_4$	$+n_5$
0	∞	2·881 944 444 444	1·118 055 555 556	·381 944 444 444	·062 500 000 000
·1	20·813 961 805 307	3·062 704 861 360	1·403 052 714 398	·506 947 285 602	·085 203 252 033
·2	10·354 662 698 413	3·235 337 301 587	1·696 884 920 635	·643 115 079 365	·110 634 920 635
·3	6·848 448 708 914	3·402 662 402 197	1·999 645 290 110	·790 354 709 890	·138 837 209 302
·4	5·084 988 540 871	3·566 678 125 796	2·311 417 112 299	·948 582 887 701	·169 848 484 848
·5	4·021 164 021 164	3·728 835 978 836	2·632 275 132 275	1·117 724 867 725	·203 703 703 704
·6	3·308 675 523 349	3·890 213 365 539	2·962 286 634 461	1·297 713 365 539	·240 434 782 609
·7	2·797 898 918 049	4·051 624 891 475	3·301 512 363 427	1·488 487 636 573	·280 070 921 986
·8	2·413 803 606 238	4·213 696 393 762	3·650 007 309 942	1·689 992 690 058	·322 638 888 889
·9	2·114 566 024 340	4·376 915 457 141	4·007 821 384 964	1·902 178 615 036	·368 163 265 306
1·0	1·875 000 000 000	4·541 666 666 667	4·375 000 000 000	2·125 000 000 000	·416 666 666 667

TABLE III (*continued*).

c_3.

λ	$-n_1$	$+n_2$	$-n_3$	$+n_4$	$-n_5$
0	∞	2·621 527 777 778	1·711 805 555 556	·663 194 444 444	·114 583 333 33[illegible]
·1	13·317 393 333 129	2·540 940 000 204	1·816 635 757 371	·738 126 147 391	·130 691 056 911
·2	6·115 845 959 596	2·475 820 707 071	1·918 623 737 374	·813 194 444 444	·147 222 222 222
·3	3·763 229 243 512	2·422 881 867 599	2·018 143 773 427	·888 377 965 704	·164 147 286 822
·4	2·616 103 259 486	2·379 730 073 848	2·115 508 021 390	·963 658 645 276	·181 439 393 939
·5	1·947 089 947 090	2·344 576 719 577	2·210 978 835 979	1·039 021 164 021	·199 074 074 074
·6	1·514 500 495 479	2·316 055 060 077	2·304 778 273 257	1·114 452 495 974	·217 028 985 507
·7	1·215 232 754 657	2·293 100 578 676	2·397 095 499 755	1·189 941 537 282	·235 283 687 943
·8	·998 046 331 106	2·274 870 335 561	2·488 092 627 402	1·265 478 801 170	·253 819 444 444
·9	·834 683 230 328	2·260 687 140 043	2·577 909 351 185	1·341 056 166 056	·272 619 047 619
1·0	·708 333 333 333	2·250 000 000 000	2·666 666 666 667	1·416 666 666 667	·291 666 666 667

c_4.

λ	$+n_1$	$-n_2$	$+n_3$	$-n_4$	$+n_5$
0	∞	·826 388 888 889	·673 611 111 111	·326 388 888 889	·062 500 000 000
·1	3·576 282 097 132	·757 051 236 202	·667 191 188 041	·332 808 811 959	·065 040 650 407
·2	1·550 099 206 349	·699 900 793 651	·661 210 317 460	·338 789 682 540	·067 460 317 460
·3	·903 489 275 582	·652 066 279 973	·655 626 027 719	·344 373 972 281	·069 767 441 860
·4	·596 829 640 947	·611 503 692 386	·650 401 069 519	·349 598 930 481	·071 969 696 970
·5	·423 280 423 280	·576 719 576 720	·645 502 645 503	·354 497 354 497	·074 074 074 074
·6	·314 512 882 448	·546 598 228 663	·640 901 771 337	·359 098 228 663	·076 086 956 522
·7	·241 614 759 762	·520 290 002 142	·636 572 742 956	·363 427 257 044	·078 014 184 397
·8	·190 363 060 429	·497 136 939 571	·632 492 690 058	·367 507 309 942	·079 861 111 111
·9	·153 007 671 805	·476 621 957 825	·628 641 200 070	·371 358 799 930	·081 632 653 061
1·0	·125 000 000 000	·458 333 333 333	·625 000 000 000	·375 000 000 000	·083 333 333 333

c_5.

λ	$-n_1$	$+n_2$	$-n_3$	$+n_4$	$-n_5$
0	∞	·086 805 555 556	·079 861 111 111	·045 138 888 889	·010 416 666 667
·1	·340 598 294 965	·076 068 371 702	·075 446 779 813	·043 600 839 234	·010 162 601 626
·2	·140 918 109 668	·067 415 223 665	·071 473 665 224	·042 162 698 413	·009 920 634 921
·3	·078 564 284 833	·060 324 604 056	·067 880 524 149	·040 815 128 024	·009 689 922 481
·4	·049 735 803 412	·054 430 863 254	·064 616 755 793	·039 549 910 873	·009 469 696 970
·5	·033 862 433 862	·049 470 899 471	·061 640 211 640	·038 359 788 360	·009 259 259 259
·6	·024 193 298 650	·045 251 145 795	·058 915 520 872	·037 238 325 282	·009 057 971 014
·7	·017 897 389 612	·041 626 419 912	·056 412 795 774	·036 179 796 818	·008 865 248 227
·8	·013 597 361 459	·038 485 971 874	·054 106 620 718	·035 179 093 567	·008 680 555 556
·9	·010 552 253 228	·035 744 043 068	·051 975 255 177	·034 231 641 375	·008 503 401 361
1·0	·008 333 333 333	·033 333 333 333	·050 000 000 000	·033 333 333 333	·008 333 333 333

TABLE IV. $\frac{1}{h_0} n_1\mu_1''$.

λ	$-n_1$	$+n_2$	$-n_3$	$+n_4$	$-n_5$
0	0	0	0	0	0
·1	·053 146 674	·000 449 173	·000 190 403	·000 067 022	·000 011 125
·2	·111 505 682	·003 389 207	·001 557 015$^+$	·000 563 833	·000 094 889
·3	·173 837 035	·010 866 342	·005 373 119	·001 998 761	·000 340 875
·4	·239 262 372	·024 618 517	·013 025 673	·004 970 771	·000 858 667
·5	·307 142 857	·046 199 157	·026 023 065$^+$	·010 174 851	·001 779 514
·6	·377 003 205$^+$	·077 054 795$^-$	·046 002 205$^+$	·018 407 333	·003 258 000
·7	·448 482 797	·118 575 634	·074 735 131	·030 571 163	·005 473 708
·8	·521 303 258	·172 129 631	·114 135 258	·047 681 123	·008 632 889
·9	·595 246 405$^+$	·239 086 220	·166 263 359	·070 869 029	·012 970 125
1·0	·670 138 889	·320 833 333	·233 333 333	·101 388 889	·018 750 000

$\frac{1}{h_0^2} n_1\mu_2''$.

λ	$+n_1$	$-n_2$	$+n_3$	$-n_4$	$+n_5$
0	0	0	0	0	0
·1	·003 650 103	·000 045 316	·000 019 321	·000 006 814	·000 001 132
·2	·015 663 867	·000 689 594	·000 320 269	·000 116 407	·000 019 625$^+$
·3	·037 282 384	·003 342 812	·001 678 446	·000 627 664	·000 107 316
·4	·069 407 157	·010 174 304	·005 487 556	·002 108 127	·000 365 345$^-$
·5	·112 711 010	·024 038 742	·013 849 817	·005 458 219	·000 958 306
·6	·167 707 836	·048 444 736	·029 670 121	·011 979 901	·002 129 814
·7	·234 798 023	·087 549 158	·056 754 081	·023 449 694	·004 219 488
·8	·314 298 975$^-$	·146 167 819	·099 910 037	·042 196 086	·007 681 354
·9	·406 466 153	·229 797 043	·165 055 118	·071 181 282	·013 103 669
1·0	·511 507 937	·344 642 857	·259 325 397	·114 087 302	·021 230 159

$\frac{1}{h_0^3} n_1\mu_3''$.

λ	$-n_1$	$+n_2$	$-n_3$	$+n_4$	$-n_5$
0	0	0	0	0	0
·1	·000 278 636	·000 004 102	·000 001 756	·000 000 620	·000 000 103
·2	·002 423 051	·000 125 549	·000 058 724	·000 021 396	·000 003 611
·3	·008 740 137	·000 917 673	·000 465 341	·000 174 614	·000 029 904
·4	·021 876 387	·003 742 508	·002 043 440	·000 788 408	·000 136 921
·5	·044 716 081	·011 104 929	·006 490 247	·002 570 858	·000 452 500$^-$
·6	·080 315 486	·026 975 571	·016 788 729	·006 818 053	·001 215 604
·7	·131 858 723	·057 118 020	·037 682 934	·015 669 565$^-$	·002 828 545$^+$
·8	·202 627 203	·109 429 483	·076 221 175$^+$	·032 414 776	·005 921 321
·9	·295 977 878	·194 300 820	·142 370 050$^+$	·061 853 498	·011 429 177
1·0	·415 327 381	·325 000 000	·249 702 381	·110 714 286	·020 684 524

TABLE IV (*continued*).

$$\frac{1}{h_0^4} n_1 \mu_4''.$$

λ	$+n_1$	$-n_2$	$+n_3$	$-n_4$	$+n_5$
0	0	0	0	0	0
·1	·000 022 553	·000 000 366	·000 000 157	·000 000 056	·000 000 009
·2	·000 395 663	·000 022 502	·000 010 577	·000 003 860	·000 000 652
·3	·002 155 292	·000 247 619	·000 126 434	·000 047 558	·000 008 154
·4	·007 232 436	·001 351 176	·000 744 069	·000 287 950	·000 050 082
·5	·018 564 080	·005 028 172	·002 967 929	·001 179 826	·000 208 029
·6	·040 168 896	·014 703 127	·009 252 508	·003 772 720	·000 674 000
·7	·077 200 853	·036 429 862	·024 325 338	·010 160 223	·001 838 150⁻
·8	·135 988 366	·079 992 630	·056 439 374	·024 117 893	·004 416 456
·9	·224 063 645⁻	·160 224 084	·119 005 540	·051 968 585⁺	·009 627 849
1·0	·350 185 185⁺	·298 558 201	·232 658 730	·103 716 931	·019 431 217

TABLE V. a_1.

λ	$-n_1$	$-n_2$	$+n_3$	$-n_4$	$+n_5$
0	∞	0	0	0	0
·1	8·174 359 079	·257 692 412	·105 943 951	·036 913 192	·006 097 561
·2	3·382 034 632	·465 367 965⁺	·201 298 701	·071 428 571	·011 904 762
·3	1·885 542 836	·635 542 836	·287 534 087	·103 770 261	·017 441 860
·4	1·193 659 282	·776 992 615⁺	·365 864 528	·134 135 472	·022 727 273
·5	·812 698 413	·896 031 746	·437 301 587	·162 698 413	·027 777 778
·6	·580 639 168	·997 305 834	·502 694 166	·189 613 527	·032 608 696
·7	·429 537 351	1·084 299 255⁺	·562 759 568	·215 018 210	·037 234 043
·8	·326 336 675⁺	1·159 670 008	·618 107 769	·239 035 088	·041 666 667
·9	·253 254 077	1·225 476 300	·669 260 542	·261 773 940	·045 918 367
1·0	·200 000 000	1·283 333 333	·716 666 667	·283 333 333	·050 000 000

a_2.

λ	$+n_1$	$-n_2$	$+n_3$	$-n_4$	$+n_5$
0	∞	5·763 888 889	+2·236 111 111	−·763 888 889	+·125 000 000
·1	34·059 829 496	4·690 170 504	+1·794 677 981	−·610 083 923	+·099 593 496
·2	14·091 810 967	3·824 855 700	+1·397 366 522	−·466 269 841	+·075 396 825⁺
·3	7·856 428 483	3·115 793 739	+1·038 052 415⁻	−·331 512 802	+·052 325 581
·4	4·973 580 341	2·526 419 659	+·711 675 579	−·204 991 087	+·030 303 030
·5	3·386 243 386	2·030 423 280	+·414 021 164	−·085 978 836	+·009 259 259
·6	2·419 329 865⁻	1·608 447 913	+·141 552 087	+·026 167 472	−·010 869 565⁺
·7	1·789 738 961	1·245 975 325⁻	−·108 720 423	+·132 020 318	−·030 141 844
·8	1·359 736 146	·931 930 521	−·339 337 928	+·232 090 643	−·048 611 111
·9	1·055 225 323	·657 737 640	−·552 474 482	+·326 835 863	−·066 326 531
1·0	·833 333 333	·416 666 667	−·750 000 000	+·416 666 667	−·083 333 333

TABLE V (*continued*).

a_3.

λ	$-n_1$	$+n_2$	$-n_3$	$+n_4$	$-n_5$
0	∞	15·729 166 667	10·270 833 333	3·979 166 667	·687 500 000
·1	71·525 641 943	13·474 358 057	9·343 823 761	3·656 176 239	·634 146 341
·2	29·592 803 030	11·657 196 970	8·509 469 697	3·354 166 667	·583 333 333
·3	16·498 499 815$^-$	10·168 166 852	7·754 910 071	3·071 176 885$^+$	·534 883 721
·4	10·444 518 717	8·930 481 283	7·069 518 717	2·805 481 283	·488 636 364
·5	7·111 111 111	7·888 888 889	6·444 444 444	2·555 555 556	·444 444 444
·6	5·080 592 716	7·002 740 617	5·872 259 383	2·320 048 309	·402 173 913
·7	3·758 451 819	6·241 548 181	5·346 687 113	2·097 757 332	·361 702 128
·8	2·855 445 906	5·582 054 094	4·862 390 351	1·887 609 649	·322 916 667
·9	2·215 973 178	5·006 249 044	4·414 803 587	1·688 644 689	·285 714 286
1·0	1·750 000 000	4·500 000 000	4·000 000 000	1·500 000 000	·250 000 000

a_4.

λ	$+n_1$	$-n_2$	$+n_3$	$-n_4$	$+n_5$
0	∞	19·833 333 333	16·166 666 667	7·833 333 333	1·500 000 000
·1	81·743 590 792	17·256 409 208	15·107 227 155$^+$	7·464 201 416	1·439 024 390
·2	33·820 346 320	15·179 653 680	14·153 679 654	7·119 047 619	1·380 952 381
·3	18·855 428 360	13·477 904 973	13·291 325 796	6·795 630 726	1·325 581 395$^+$
·4	11·936 592 819	12·063 407 181	12·508 021 390	6·491 978 610	1·272 727 273
·5	8·126 984 127	10·873 015 873	11·793 650 794	6·206 349 206	1·222 222 222
·6	5·806 391 676	9·860 274 991	11·139 725 009	5·937 198 068	1·173 913 043
·7	4·295 373 507	8·990 340 779	10·539 070 986	5·683 151 236	1·127 659 574
·8	3·263 366 750$^+$	8·236 633 250$^-$	9·985 588 972	5·444 982 456	1·083 333 333
·9	2·532 540 775$^-$	7·578 570 336	9·474 061 243	5·215 593 930	1·040 816 327
1·0	2·000 000 000	7·000 000 000	9·000 000 000	5·000 000 000	1·000 000 000

a_5.

λ	$-n_1$	$+n_2$	$-n_3$	$+n_4$	$-n_5$
0	∞	10·416 666 667	9·583 333 333	5·416 666 667	1·250 000 000
·1	40·871 795 396	9·128 204 604	9·053 613 578	5·232 100 708	1·219 512 195$^+$
·2	16·910 173 160	8·089 826 840	8·576 839 827	5·059 523 810	1·190 476 190
·3	9·427 714 180	7·238 952 487	8·145 662 898	4·897 815 363	1·162 790 698
·4	5·968 296 409	6·531 703 591	7·754 010 695$^+$	4·745 989 305$^-$	1·136 363 636
·5	4·063 492 063	5·936 507 937	7·396 825 397	4·603 174 603	1·111 111 111
·6	2·903 195 838	5·430 137 495$^+$	7·069 862 505$^-$	4·468 599 034	1·086 956 522
·7	2·147 686 753	4·995 170 389	6·769 535 493	4·341 575 618	1·063 829 787
·8	1·631 683 375$^+$	4·618 316 625$^-$	6·492 794 486	4·221 491 228	1·041 666 667
·9	1·266 270 387	4·289 285 168	6·237 030 621	4·107 796 965$^-$	1·020 408 163
1·0	1·000 000 000	4·000 000 000	6·000 000 000	4·000 000 000	1·000 000 000

TABLE VI. K_1.

λ	n_1	n_2	n_3	n_4	n_5
0	$-\infty$	$-\cdot021\,501\,598$	$+\cdot013\,948\,137$	$-\cdot005\,347\,498$	$+\cdot000\,913\,525^{+}$
·1	$-\cdot636\,353\,675^{+}$	$-\cdot039\,437\,722$	$+\cdot021\,316\,401$	$-\cdot007\,914\,081$	$+\cdot001\,337\,436$
·2	$-\cdot352\,799\,985^{-}$	$-\cdot051\,314\,487$	$+\cdot026\,752\,959$	$-\cdot009\,879\,800$	$+\cdot001\,667\,992$
·3	$-\cdot308\,362\,782$	$-\cdot055\,978\,631$	$+\cdot029\,089\,396$	$-\cdot010\,752\,319$	$+\cdot001\,817\,055^{+}$
·4	$-\cdot324\,425\,067$	$-\cdot052\,318\,318$	$+\cdot027\,025\,398$	$-\cdot009\,946\,742$	$+\cdot001\,676\,356$
·5	$-\cdot365\,125\,556$	$-\cdot039\,230\,633$	$+\cdot019\,124\,747$	$-\cdot006\,780\,511$	$+\cdot001\,115\,842$
·6	$-\cdot418\,429\,429$	$-\cdot015\,600\,485^{+}$	$+\cdot003\,811\,099$	$-\cdot000\,468\,312$	$-\cdot000\,017\,978$
·7	$-\cdot479\,128\,525^{+}$	$+\cdot019\,713\,730$	$-\cdot020\,636\,406$	$+\cdot009\,882\,997$	$-\cdot001\,903\,687$
·8	$-\cdot544\,586\,042$	$+\cdot067\,890\,333$	$-\cdot056\,087\,666$	$+\cdot025\,279\,452$	$-\cdot004\,746\,618$
·9	$-\cdot613\,315\,045^{+}$	$+\cdot130\,151\,913$	$-\cdot104\,566\,227$	$+\cdot046\,845\,034$	$-\cdot008\,780\,513$
1·0	$-\cdot684\,408\,069$	$+\cdot207\,771\,164$	$-\cdot168\,253\,968$	$+\cdot075\,826\,720$	$-\cdot014\,269\,180$

K_2.

λ	n_1	n_2	n_3	n_4	n_5
0	$-\infty$	$+\cdot041\,473\,765^{+}$	$-\cdot013\,288\,139$	$+\cdot003\,775\,353$	$-\cdot000\,545\,635^{-}$
·1	$-\cdot253\,150\,199$	$+\cdot033\,332\,954$	$-\cdot009\,940\,553$	$+\cdot002\,608\,898$	$-\cdot000\,352\,946$
·2	$-\cdot090\,583\,915^{-}$	$+\cdot026\,164\,477$	$-\cdot006\,644\,003$	$+\cdot001\,414\,993$	$-\cdot000\,152\,018$
·3	$-\cdot021\,952\,593$	$+\cdot018\,165\,156$	$-\cdot002\,576\,711$	$-\cdot000\,112\,290$	$+\cdot000\,109\,623$
·4	$+\cdot031\,907\,940$	$+\cdot006\,889\,972$	$+\cdot003\,693\,177$	$-\cdot002\,546\,686$	$+\cdot000\,533\,695^{+}$
·5	$+\cdot087\,179\,810$	$-\cdot010\,714\,122$	$+\cdot014\,299\,657$	$-\cdot006\,794\,093$	$+\cdot001\,285\,319$
·6	$+\cdot149\,466\,856(5)$	$-\cdot038\,301\,676$	$+\cdot032\,174\,292$	$-\cdot014\,161\,322$	$+\cdot002\,608\,592$
·7	$+\cdot221\,303\,959$	$-\cdot080\,139\,026$	$+\cdot061\,145\,227$	$-\cdot026\,429\,212$	$+\cdot004\,843\,573$
·8	$+\cdot304\,046\,996$	$-\cdot141\,125\,485^{+}$	$+\cdot106\,039\,965(5)$	$-\cdot045\,930\,103$	$+\cdot008\,444\,692$
·9	$+\cdot398\,510\,089$	$-\cdot226\,822\,037$	$+\cdot172\,792\,029$	$-\cdot075\,629\,647$	$+\cdot014\,000\,575^{+}$
1·0	$+\cdot505\,224\,868$	$-\cdot343\,485\,450^{-}$	$+\cdot268\,551\,587$	$-\cdot119\,212\,963$	$+\cdot022\,255\,291$

K_3.

λ	n_1	n_2	n_3	n_4	n_5
0	$+\infty$	$+\cdot030\,123\,595^{-}$	$-\cdot019\,380\,374$	$+\cdot007\,330\,936$	$-\cdot001\,233\,879$
·1	$+\cdot066\,421\,864$	$+\cdot032\,230\,396$	$-\cdot020\,246\,603$	$+\cdot007\,632\,757$	$-\cdot001\,283\,736$
·2	$+\cdot025\,173\,412$	$+\cdot034\,046\,417$	$-\cdot021\,081\,639$	$+\cdot007\,935\,169$	$-\cdot001\,334\,630$
·3	$+\cdot006\,645\,369$	$+\cdot036\,227\,120$	$-\cdot022\,191\,913$	$+\cdot008\,352\,286$	$-\cdot001\,406\,104$
·4	$-\cdot012\,136\,459$	$+\cdot040\,206\,147$	$-\cdot024\,409\,167$	$+\cdot009\,213\,852$	$-\cdot001\,556\,248$
·5	$-\cdot038\,084\,688$	$+\cdot048\,539\,894$	$-\cdot029\,438\,880$	$+\cdot011\,229\,368$	$-\cdot001\,913\,038$
·6	$-\cdot075\,577\,632$	$+\cdot065\,236\,904$	$-\cdot040\,270\,947$	$+\cdot015\,696\,183$	$-\cdot002\,715\,561$
·7	$-\cdot128\,353\,817$	$+\cdot096\,089\,196$	$-\cdot061\,655\,270$	$+\cdot024\,754\,989$	$-\cdot004\,366\,244$
·8	$-\cdot199\,964\,386$	$+\cdot149\,015\,662(5)$	$-\cdot100\,645\,137$	$+\cdot041\,696\,171$	$-\cdot007\,495\,188$
·9	$-\cdot293\,911\,395^{+}$	$+\cdot234\,423\,961$	$-\cdot167\,211\,404$	$+\cdot071\,320\,436$	$-\cdot013\,037\,737$
1·0	$-\cdot413\,695\,437$	$+\,365\,595\,238$	$-\cdot274\,930\,556$	$+\cdot120\,357\,143$	$-\cdot022\,326\,389$

K_4.

λ	n_1	n_2	n_3	n_4	n_5
0	$+\infty$	$-\cdot031\,972\,002$	$+\cdot006\,540\,062$	$-\cdot000\,622\,795^{+}$	$-\cdot000\,049\,603$
·1	$+\cdot213\,572\,278$	$-\cdot025\,240\,324$	$+\cdot003\,752\,503$	$+\cdot000\,341\,482$	$-\cdot000\,208\,889$
·2	$+\cdot088\,749\,081$	$-\cdot019\,837\,073$	$+\cdot001\,271\,844$	$+\cdot001\,239\,368$	$-\cdot000\,359\,955^{+}$
·3	$+\cdot051\,413\,852$	$-\cdot015\,616\,484$	$+\cdot000\,865\,142$	$+\cdot002\,040\,576$	$-\cdot000\,497\,106$
·4	$+\cdot038\,415\,996$	$-\cdot013\,024\,759$	$-\cdot002\,293\,838$	$+\cdot002\,593\,455^{+}$	$-\cdot000\,593\,256$
·5	$+\cdot039\,795\,288$	$-\cdot013\,591\,937$	$-\cdot001\,936\,224$	$+\cdot002\,447\,767$	$-\cdot000\,567\,250^{+}$
·6	$+\cdot055\,337\,711$	$-\cdot020\,621\,173$	$+\cdot002\,640\,017$	$+\cdot000\,558\,013$	$-\cdot000\,227\,483$
·7	$+\cdot088\,422\,233$	$-\cdot040\,075\,263$	$+\cdot016\,143\,678$	$-\cdot005\,165\,810$	$+\cdot000\,815\,832$
·8	$+\cdot144\,513\,696$	$-\cdot081\,669\,020$	$+\cdot046\,811\,780$	$-\cdot018\,496\,055^{-}$	$+\cdot003\,278\,339$
·9	$+\cdot230\,679\,740$	$-\cdot160\,181\,328$	$+\cdot108\,041\,613$	$-\cdot045\,752\,710$	$+\cdot008\,378\,659$
1·0	$+\cdot355\,410\,053$	$-\cdot297\,003\,968$	$+\cdot220\,456\,349$	$-\cdot096\,937\,831$	$+\cdot018\,075\,397$

ON ASYMPTOTIC FORMULAE FOR THE HYPERGEOMETRIC SERIES

II. HYPERGEOMETRIC SERIES IN WHICH THE FOURTH ELEMENT, x, IS NOT NECESSARILY UNITY

By O. L. DAVIES, Ph.D.

The hypergeometric series was first introduced by Euler in 1778 when attempting to obtain solutions of the second order differential equation

$$x^n(a+bx^n)\frac{d^2y}{dx^2}+x(c+ex^n)\frac{dy}{dx}+(f+gx^n)y=0.$$

A particular case of this equation is

$$x(1-x)\frac{d^2y}{dx^2}+\{\gamma-(\alpha+\beta+1)x\}\frac{dy}{dx}-\alpha\beta y=0,$$

which has the hypergeometric function

$$F(\alpha,\beta,\gamma,x)=1+\frac{\alpha\beta}{1!\gamma}x+\frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\gamma(\gamma+1)}x^2+\ldots$$

as one of its solutions.

The chief problem in the study of series is to determine the partial sums and in particular, the total sum. In this paper we will obtain approximations to the sum of a number of terms of the hypergeometric series by fitting certain systems of frequency curves to the series and then expressing, approximately, the required sum by means of an integral.

The hypergeometric series will be looked upon as a discrete frequency distribution; we restrict our attention, therefore, to series with positive terms only.

In § 1 the system of frequency curves which arises from the differential equation

$$\frac{1}{y}\frac{dy}{dz}=\frac{y_{r+1}-y_r}{\frac{c}{2}(y_{r+1}+y_r)},$$

where y_r is the rth term of the series, is discussed. This equation was originally employed by Professor Karl Pearson to obtain curves to fit the series $F(\alpha, \beta, \gamma, 1)$ in which the last element is unity. The curves which arise from its integration are known as the Pearson System of Frequency Curves. These will be denoted in the subsequent work by $P(z)$. When the last element of F is not unity, the resulting system of frequency curves has the more general form $e^{-pz}P(z)$. The goodness of fit of this new set of curves to the general series is tested in a variety of examples.

In § 2 an alternative method of curve fitting is given.

§§ 3—6 are devoted to the evaluation of the integral $\int_{z_1}^{z_2} e^{-\rho z} P(z)\, dz$.

§§ 7—9 give further and more accurate methods for summing the tail of the series, while § 10 is mainly a summary of the results and conclusions arrived at in the previous paragraphs.

§ 1. *On the Fitting of a System of Frequency Curves to the Hypergeometric Series* $F(\alpha, \beta, \gamma, x)$.

Let y_r denote the rth term of the hypergeometric series

$$1+\frac{\alpha\beta}{1!\gamma}x+\frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\gamma(\gamma+1)}x^2+\ldots,$$

$$y_r=\frac{\alpha(\alpha+1)\ldots(\alpha+r-2)\beta(\beta+1)\ldots(\beta+r-2)}{(r-1)!\gamma(\gamma+1)\ldots(\gamma+r-2)}x^{r-1}.$$

Form the ratio

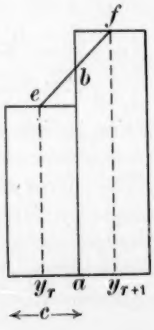
$$\frac{y_{r+1}-y_r}{y_{r+1}+y_r}=\frac{\frac{(\alpha+r-1)(\beta+r-1)}{r(\gamma+r-1)}x-1}{\frac{(\alpha+r-1)(\beta+r-1)}{r(\gamma+r-1)}x+1}$$

$$=\frac{r^2(x-1)+r(\overline{\alpha+\beta-2}x-\overline{\gamma-1})+(\alpha-1)(\beta-1)x}{r^2(x+1)+r(\overline{\alpha+\beta-2}x+\overline{\gamma-1})+(\alpha-1)(\beta-1)x}.$$

Let each term of the hypergeometric series be represented proportionally by a rectangular block of width c and assume the weight of each block to be concentrated at the mid-vertical. The ratio

$$\frac{\Delta y_r}{\frac{c}{2}(y_r+y_{r+1})}$$

is equal to the slope of the line ef divided by ab, ab being the ordinate of ef at the point a.

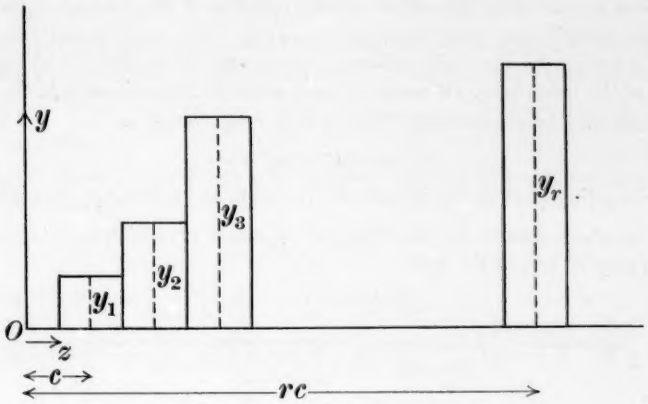


We wish to test the goodness of fit of the curves for which $\frac{1}{y}\frac{dy}{dz}$ is equal to this ratio at all points a midway between the mid-verticals of the rectangular blocks composing the histogram of the series*. Evidently, the equation of the required curve is

$$\frac{1}{y}\frac{dy}{dz} = \frac{\Delta y_r}{\frac{c}{2}(y_r + y_{r+1})} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1),$$

where $rc = z$, measured from a suitable origin.

The distance of the rth mid-ordinate from a point O, distance c before the first mid-ordinate, is rc.



The relation (1) is true for all points at distances $(r+\frac{1}{2})c$, $(r = 0, 1, 2, \ldots)$ from O. Hence the equation of the curve which fits the series is

$$\frac{1}{y}\frac{dy}{dz} = \left\{\frac{\Delta y_r}{c\left(\frac{y_r + y_{r+1}}{2}\right)}\right\}_{(r+\frac{1}{2})c = z},$$

z being measured from the origin O.

It will simplify matters considerably if the origin be moved to the start of the histogram. This necessitates replacing rc of (1) by z. The equation of the required curve referred to the new origin is then

$$\frac{1}{y}\frac{dy}{dz} = \frac{2}{c}\,\frac{z^2(x-1) + zc(\overline{\alpha+\beta-2}\,x - \overline{\gamma-1}) + (\alpha-1)(\beta-1)xc^2}{z^2(x+1) + zc(\overline{\alpha+\beta-2}\,x + \overline{\gamma-1}) + (\alpha-1)(\beta-1)xc^2} \quad \ldots\ldots (2).$$

The mid-verticals of the rectangular blocks are now at the points

$$z = (r+\tfrac{1}{2})c, \quad (r = 0, 1, 2, 3, \ldots).$$

* The reader must bear in mind that we are *not* proceeding to a limit; c is finite.

The terms of the hypergeometric series will continue to increase as long as $y_{r+1}/y_r > 1$. It is reasonable, therefore, to determine the position of the modes from the values of r which satisfy $y_{r+1}/y_r = 1$, or what is equivalent,

$$\Delta y_r = 0 \quad \ldots\ldots\ldots\

Then $$\frac{1}{y}\frac{dy}{dz}=\frac{2}{c}\left[\left(\frac{x-1}{x+1}\right)+\frac{2xc}{(x+1)^2}\left\{\frac{z(\alpha+\beta-\gamma-1)+(\alpha-1)(\beta-1)c}{(z+F_1)(z+F_2)}\right\}\right]$$

$$=\frac{2}{c}\left(\frac{x-1}{x+1}\right)+\frac{4x}{(x+1)^2}\left\{\frac{A}{z+F_1}+\frac{B}{z+F_2}\right\},$$

where $$2A=-(\epsilon+2)-\frac{1}{\sqrt{\Delta}}\{2(x+1)(\alpha\beta-\gamma)+(\epsilon+2)(\overline{\alpha+\beta}.x+\overline{\gamma+1})\},$$

$$2B=-(\epsilon+2)+\frac{1}{\sqrt{\Delta}}\{2(x+1)(\alpha\beta-\gamma)+(\epsilon+2)(\overline{\alpha+\beta}.x+\overline{\gamma+1})\}$$

and $$\epsilon=\gamma-\alpha-\beta-1.$$

Integrating the original differential equation, we find

$$y=y_0 e^{-\frac{2}{c}\left(\frac{1-x}{1+x}\right)z}[z+F_1]^{\frac{4Ax}{(x+1)^2}}[z+F_2]^{\frac{4Bx}{(x+1)^2}}.$$

This is a product of an exponential term into a Pearson Type I or VI curve, the former when F_1 and F_2 are of opposite sign and the latter when F_1 and F_2 are of the same sign.

II. Δ *negative.*

The denominator of the differential equation can now be expressed as a sum of two squares, namely,

$$\left[z+\frac{c}{2(x+1)}\{(\alpha+\beta-2)x+(\gamma-1)\}\right]^2+\frac{c^2D}{4(x+1)^2},$$

where $D=|\Delta|$. Then

$$\frac{1}{y}\frac{dy}{dz}=-\frac{2}{c}\left(\frac{1-x}{1+x}\right)+\frac{4x}{(x+1)^2}\cdot\frac{z(\alpha+\beta-\gamma-1)+(\alpha-1)(\beta-1)c}{R^2+\frac{c^2D}{4(x+1)^2}},$$

where $$R=\left[z+\frac{c}{2(x+1)}\{(\alpha+\beta-2)x+(\gamma-1)\}\right].$$

Let $\zeta=R\dfrac{2(x+1)}{c\sqrt{D}}$, then

$$\frac{1}{y}\frac{dy}{d\zeta}=-\frac{\sqrt{D}(1-x)}{(1+x)^2}-\frac{2x(\epsilon+2)}{(1+x)^2}\cdot\frac{2\zeta}{1+\zeta^2}+\frac{\nu}{1+\zeta^2},$$

where $$\nu=\frac{8x}{\sqrt{D}(x+1)}\left[(\alpha-1)(\beta-1)+\frac{\epsilon+2}{2(x+1)}\{(\alpha+\beta-2)x+(\gamma-1)\}\right].$$

Hence, on integration,

$$y=y_0 e^{-\sqrt{D}\left\{\frac{1-x}{(1+x)^2}\right\}\zeta}(1+\zeta^2)^{-2x\cdot\frac{\epsilon+2}{(x+1)^2}}e^{\nu\arctan\zeta}$$

$$=y_0' e^{-\frac{2}{c}\left(\frac{1-x}{1+x}\right)z}\frac{e^{\nu\arctan\frac{2(x+1)}{c\sqrt{D}}R}}{\left[R^2+\frac{c^2D}{4(x+1)^2}\right]^{2x\frac{(\epsilon+2)}{(x+1)^2}}}.$$

This is a product of an exponential term into a Pearson Type IV curve.

III. $\Delta = 0$.

For this case

$$\frac{1}{y}\frac{dy}{dz} = \frac{2(x-1)}{c(x+1)} + \frac{4x}{(x+1)^2}\cdot\frac{z(\alpha+\beta-\gamma-1)+(\alpha-1)(\beta-1)c}{\left[z+\frac{c}{2(x+1)}\{(\alpha+\beta-2)x+(\gamma-1)\}\right]^2}$$

and we readily find

$$y = y_0 e^{-\frac{2}{c}\left(\frac{1-x}{1+x}\right)z}\left[z+\frac{c}{2(x+1)}\{(\alpha+\beta-2)x+(\gamma-1)\}\right]^{-\frac{4x(\epsilon+2)}{(x+1)^2}} \times e^{-\frac{4xcv'}{(x+1)^2}\Big/\left[z+\frac{c}{2(x+1)}\{(\alpha+\beta-2)x+(\gamma-1)\}\right]},$$

where $\nu' = [(\alpha-1)(\beta-1)+\{(\alpha+\beta-2)x+(\gamma-1)\}(\epsilon+2)/2(x+1)]$.

This is a product of an exponential term into a Pearson Type V curve.

IV. *Type III Curves.*

When either α or β is unity, the differential equation reduces to the fairly simple form

$$\frac{1}{y}\frac{dy}{dz} = \frac{2}{c}\frac{z(x-1)+c(\overline{\alpha+\beta-2}x-\overline{\gamma-1})}{z(x+1)+c(\overline{\alpha+\beta-2}x+\overline{\gamma-1})}$$

$$= -\frac{2}{c}\left(\frac{1-x}{1+x}\right) - \frac{4x(\epsilon+2)}{(x+1)^2}\cdot\frac{1}{z+\frac{c}{x+1}(\overline{\alpha+\beta-2}x+\overline{\gamma-1})} \quad \ldots\text{(i)},$$

which gives the solution

$$y = y_0 e^{-\frac{2}{c}\left(\frac{1-x}{1+x}\right)z}\left[z+\frac{c}{x+1}(\overline{\alpha+\beta-2}x+\overline{\gamma-1})\right]^{-4x\frac{\epsilon+2}{(x+1)^2}} \quad \ldots\ldots\text{(ii)}.$$

Since the hypergeometric series is symmetrical with respect to α and β, we may take $\beta = 1$. (ii) then reduces to the final form:

$$y = y_0 e^{-\frac{2}{c}\left(\frac{1-x}{1+x}\right)z}\left[z+\frac{c}{x+1}\{(\alpha-1)x+(\gamma-1)\}\right]^{-4x\frac{(\gamma-\alpha)}{(x+1)^2}}, \qquad 0 \leqslant z \leqslant \infty.$$

This is a Type III curve when the quantity $\psi = c\dfrac{(\alpha-1)x+(\gamma-1)}{x+1}$ is positive. When negative, the solution of (i) is

$$y = y_0' e^{-\frac{2}{c}\left(\frac{1-x}{1+x}\right)z}[\psi'-z]^{-4x\frac{(\gamma-\alpha)}{(x+1)^2}}, \qquad 0 \leqslant z \leqslant \psi',$$

where $\psi' = |\psi|$. This curve is not of Type III because its range is limited. It may be looked upon as a product of an exponential term into a Pearson Type VIII or IX curve, which we may symbolise as G_{VIII} or G_{IX}.

When the range of the series extends beyond ψ', as would be the case were α and γ positive, fractional and less than unity, the equation of the curve corresponding to terms of the series beyond ψ' is

$$y = y_0'' e^{-\frac{2}{c}\left(\frac{1-x}{1+x}\right)z}[z-\psi']^{-4x\frac{(\gamma-\alpha)}{(x+1)^2}}, \qquad \psi' \leqslant z \leqslant \infty.$$

Hence, the complete solution of (i) when ψ is negative is the following:

$$y = y_0' e^{-\frac{2}{c}\left(\frac{1-x}{1+x}\right)z} [\psi' - z]^{-4x\frac{(\gamma-\alpha)}{(x+1)^2}}, \qquad 0 \leqslant z \leqslant \psi',$$

$$= y_0'' e^{-\frac{2}{c}\left(\frac{1-x}{1+x}\right)z} [z - \psi']^{-4x\frac{(\gamma-\alpha)}{(x+1)^2}}, \qquad z > \psi'.$$

When $z = \psi'$, y becomes zero. We cannot, therefore, expect a good fit near this point.

When α and γ are negative and fairly large numbers, the terms of the series beyond the point ψ' are small and generally negligible compared with the modal terms. For such cases the second part of the solution of the differential equation will not be required for computing purposes.

Sufficient conditions for a Type III curve are, therefore,

(*a*) α or β or both unity,

(*b*) $[(\alpha-1)x + (\gamma-1)]/(x+1)$ positive (for $\beta = 1$).

A necessary condition is that the numerator and denominator of the fundamental differential equation have a factor in common.

The curves which arise out of the solution of the differential equation

$$\frac{1}{y}\frac{dy}{dz} = \frac{az^2 + 2bz + d}{Az^2 + 2Bz + D}$$

have the form $$y = e^{-\rho z} P(z),$$

where $P(z)$ is a Pearson type curve. These may, therefore, be divided into three main types, namely,

G_{I}: $\quad y = y_0 e^{-\rho z} z^r (a-z)^s, \qquad 0 \leqslant z \leqslant a,$

G_{VI}: $\quad y = y_0 e^{-\rho z} (z-a)^p z^{-q}, \qquad a \leqslant z \leqslant \infty,$

G_{IV}: $\quad y = y_0 e^{-\rho z} \dfrac{e^{-\nu \arctan z/a}}{(1 + z^2/a^2)^r}, \qquad -\infty \leqslant z \leqslant +\infty.$

Integrating throughout the range for curves of type G_{I}, we have

$$N = y_0 \int_0^a e^{-\rho z} z^r (a-z)^s \, dz$$

$$= y_0 \sum_0^\infty \frac{(-)^t}{t!} \rho^t \int_0^a z^{r+t} (a-z)^s \, dz$$

$$= y_0 a^{r+s+1} \frac{\Gamma(s+1)\Gamma(r+1)}{\Gamma(r+s+2)} \left[1 - \frac{r+1}{1!(r+s+2)}\rho a + \frac{(r+1)(r+2)}{2!(r+s+2)(r+s+3)}\rho^2 a^2 - \ldots\right].$$

The higher moments are given by the partial differentials of N with respect to ρ, i.e.

$$N\mu_s' = (-)^s \frac{\partial^s N}{\partial \rho^s}.$$

The moments of the curves G_{VI} and G_{IV} may similarly be expressed by infinite series. I have attempted to use these series in a number of numerical examples but have found that they converge far too slowly to be of practical use.

Examples of the Above Method of Curve Fitting.

I. *Series F* (30, 30, 111, ·75) *with* $c = 1$.

The differential equation of the curve for which $d \log y/dz$ is equal to

$$\Delta y_r / \tfrac{1}{2}(y_{r+1} + y_r), \quad r = 0, 1, 2, \ldots,$$

y_r being the rth term of the series, is the following:

$$\frac{1}{y}\frac{dy}{dz} = -\frac{2}{7} + \frac{13{\cdot}203{,}4565}{z + 4{\cdot}322{,}090} - \frac{64{\cdot}142{,}2321}{z + 83{\cdot}392{,}196}.$$

Its solution is

$$y = y_0 e^{-\frac{3}{7}z} (z + 4{\cdot}322{,}090)^{13{\cdot}203{,}457} (z + 83{\cdot}392{,}196)^{-64{\cdot}142{,}232}.$$

This curve belongs to type G_{IV}.

The equation giving the position of the modes of the curve is

$$z^2(x-1) + z\{(\alpha+\beta-2)x - (\gamma-1)\} + (\alpha-1)(\beta-1)x = 0.$$

This equation also gives the position of the modes of the series. It has two roots, namely, 9·168,914; $-275{\cdot}16891$. The latter is well outside the range of the series and must accordingly be rejected. Hence, the true position of the mode is 9·168,914 (i.e. the tenth term) *measured from the start of the histogram.*

The ordinates of the curve corresponding to the mid-points of the rectangular blocks representing the terms of the series are obtained by putting

$$z = r + \tfrac{1}{2} \quad (r = 0, 1, 2, 3, \ldots)$$

in y.

The constant y_0 may be found by equating the value of one of the mid-ordinates of the curve to the corresponding term of the series. The best results will be obtained by selecting the mid-ordinate nearest a point of inflexion of the curve because here the area under the curve corresponds most closely to the numerical value of the mid-ordinate. This method has the objection of throwing too much weight on a single ordinate. This objection will shortly be removed when other methods of curve fitting are discussed (§ 2).

When we require merely the probability integral of the series (or the curve) and not the actual value of the partial sums (or partial areas), the question as to the choice of y_0 does not arise, because a probability integral measures the

proportional values of the frequencies and not their absolute values. In most of the subsequent examples the constant y_0 is calculated by equating the whole area of the curve (found by quadrature) to the total sum of the series (found mechanically). This is done to compare the *Probability Integral* of the series with the *Probability Integral* of the generalised Pearson curve which fits it. We do not know the actual sum of the general hypergeometric series and no one has yet succeeded in finding a simple and accurate expression for it. One method of obtaining an approximation to this sum is to calculate y_0 by equating the ordinate of the curve at a selected point to the corresponding term of the series. How good this approximation is may be judged from the subsequent examples.

In this first example, y_0 is calculated by equating the ordinates at a selected point.

Terms of Series	Mid-ordinates of Curve	Areas under Curve	Terms of Series	Mid-ordinates of Curve	Areas under Curve
1·000	1·702	1·866	58·705	59·024	59·176
6·081	7·202	7·539	43·152	43·496	43·638
19·567	20·703	21·166	31·030	31·373	31·496
44·397	45·067	45·503	21·877	22·198	22·299
79·328	79·595	79·826	15·152	15·438	15·517
119·717	119·496	119·421	10·327	10·572	10·632
158·031	157·618	157·252	6·937	7·139	7·183
187·553	187·124	186·573	4·599	4·760	4·937
203·994	203·639	203·042	3·013	3·138	3·160
206·280	206·015	205·490	1·952	2·047	2·062
196·095	195·916	195·535	1·252	1·324	1·334
176·795	176·702	176·484	·796	·847	·896
152·250	152·250	152·178	·502	·539	·544
125·971	126·068	126·107	·314	·340	·343
100·628	100·818	100·928	·195	·213	·215
77·926	78·198	78·342	·121	·134	·135
			...	...	...

II. *Series* $F(-30, -50, 100, \cdot 5)$; $c = 1$.

The equation of the curve which fits this series is

$$y = y_0 e^{-\frac{3}{2}z} \frac{e^{\frac{121928}{9\sqrt{(1379)}} \arctan \frac{3R}{\sqrt{(1379)}}}}{\left(R^2 + \frac{1379}{9}\right)^{\frac{724}{9}}},$$

where $$R = z + \tfrac{58}{3}.$$

$$R_{(\text{mode})} = 24{\cdot}870{,}2702.$$

This curve belongs to G_{IV}.

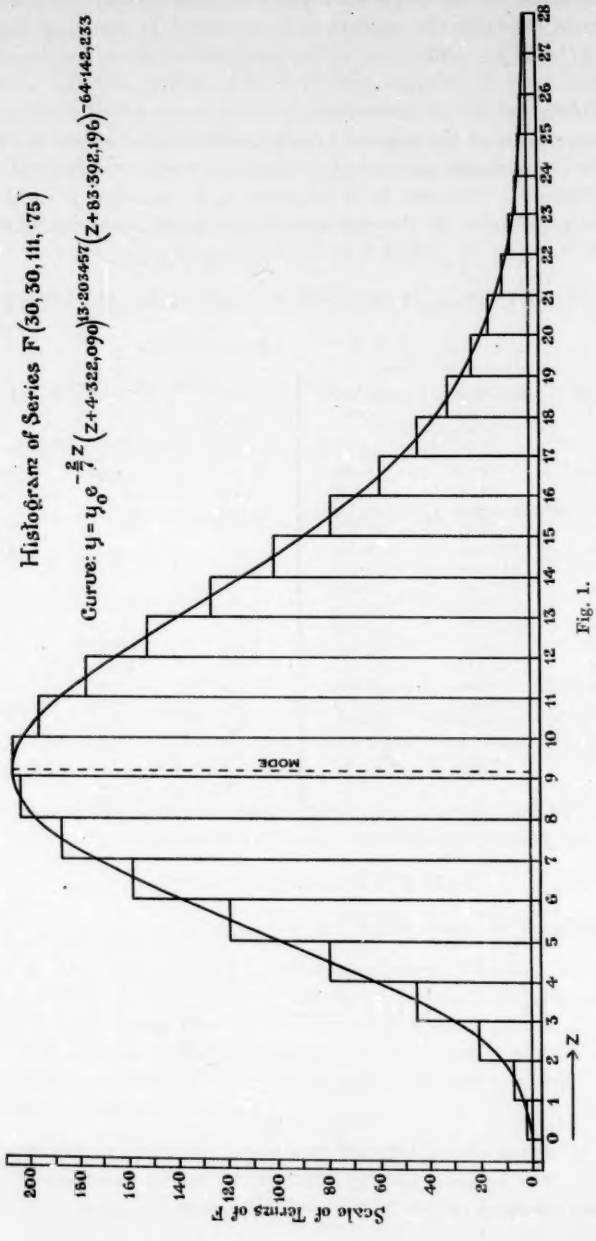

Histogram of Series F(30, 30, 111, ·75)
Curve: $y = y_0 e^{-\frac{2}{7}z}(z+4{\cdot}322{,}090)^{13{\cdot}203457}(z+83{\cdot}392{,}196)^{-64{\cdot}142{,}233}$
MODE
Scale of Terms of F
0 20 40 60 80 100 120 140 160 180 200
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
z

Fig. 1.

In this example, and in all subsequent ones, the constant y_0 is determined by equating the area of the curve, found by quadrature, to the sum of the series.

Terms of Series	Mid-ordinates of Curve	Areas under Curve	Terms of Series	Mid-ordinates of Curve	Areas under Curve
1·000	1·911	2·124	37·658	37·766	38·098
7·500	8·840	9·324	17·899	18·558	18·921
26·380	27·000	27·538	7·069	7·800	8·043
57·932	57·173	57·201	2·337	2·854	2·976
89·219	87·608	86·861	·650	·923	·974
102·602	100·864	99·994	·153	·268	·285
91·609	90·179	89·536	·030	·071	·076
65·188	64·446	64·423	·005	·016	·018
			...	...	...

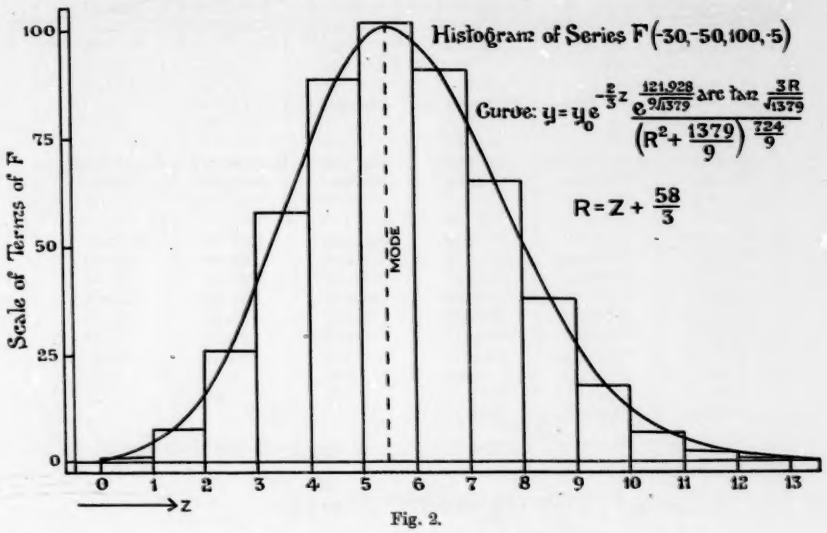


Fig. 2.

III. *Series* $F(-30, 60, -81, \frac{1}{3})$; $c = 1$.

The equation of the curve which fits this series is

$$y = y_0 e^{-z} (z + 7{\cdot}388{,}310)^{28{\cdot}509{,}628} (61{\cdot}888{,}310 - z)^{53{\cdot}900{,}372}$$

which belongs to type G_1.

$$z_{(\text{mode})} = 7{\cdot}036{,}596.$$

Terms of Series	Mid-ordinates of Curve	Areas under Curve	Terms of Series	Mid-ordinates of Curve	Areas under Curve
1·000	1·993	2·318	92·107	92·054	92·405
7·407	11·184	11·512	55·036	55·637	56·115
27·299	29·002	29·901	29·462	30·381	30·803
66·655	67·169	67·817	14·189	15·094	15·388
121·133	120·184	120·174	6·165	6·863	7·040
174·515	172·667	172·412	2·421	2·870	2·963
207·300	205·218	203·876	·860	1·108	1·152
208·485	206·550	205·333	·276	·397	·416
180·898	179·374	178·696	·080	·131	·138
137·303	136·452	136·405	...	...	...

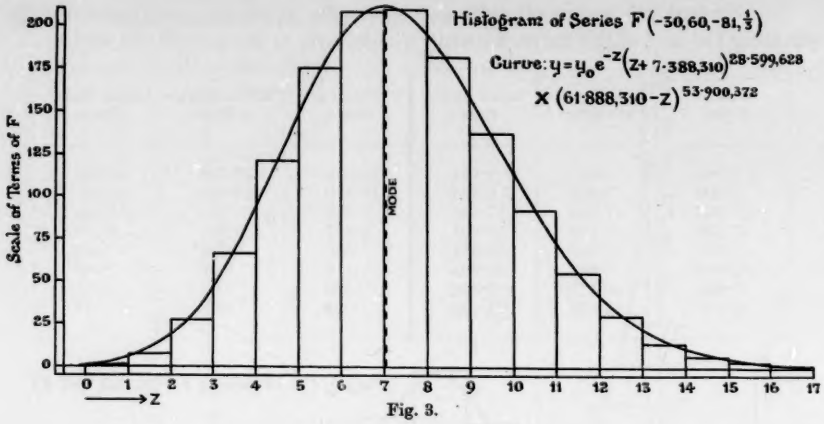


Fig. 3.

IV. *Series* $F(-20, 20, -200, 5)$; $c=1$.

Terms of Series	Mid-ordinates of Curve	Areas under Curve	Terms of Series	Mid-ordinates of Curve	Areas under Curve
1·00	3·14	3·53	1609·69	1595·46	1589·87
10·00	16·185	17·48	1155·28	1148·56	1149·18
50·13	61·59	64·67	710·59	713·07	717·61
167·09	181·39	186·50	372·16	381·60	386·96
414·53	424·38	430·63	164·19	175·69	179·89
812·15	812·09	816·12	60·03	69·39	71·88
1301·52	1291·04	1289·33	17·74	23·38	24·56
1744·30	1727·35	1718·82	4·08	6·68	7·17
1982·67	1963·62	1951·16	·69	1·61	1·76
1927·60	1909·41	1898·31			

The curve which fits this series is

$$y = y_0 e^{\frac{4}{3}z}(z+7{\cdot}747{,}944)^{21{\cdot}282{,}877}(42{\cdot}914{,}611-z)^{89{\cdot}272{,}679},$$

$$z_{(\text{mode})} = 8{\cdot}816{,}981.$$

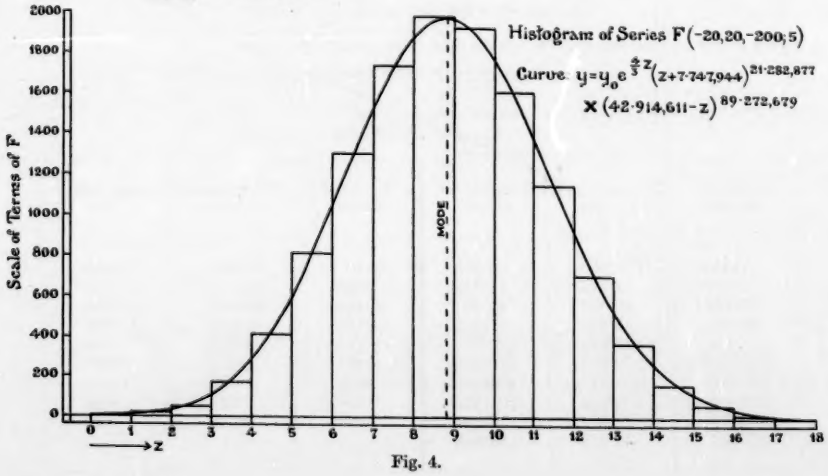


Fig. 4.

V. *Series* $F(1, -60, -65, \cdot 9)$; $c = 1$.

Equation of the curve which fits the above series is

$$y = y_0\, e^{-105{,}263z}\,(63{\cdot}631{,}579 - z)^{47{\cdot}366{,}150}.$$

This curve belongs to type G_{IX} (see p. 64).

Terms of Series	Mid-ordinates of Curve	Areas under Curve	Terms of Series	Mid-ordinates of Curve	Areas under Curve
1·000,000	·993,952	·995,309	·003,534	·003,586	·003,595
·830,769	·826,180	·827,323	·002,762	·002,806	·002,813
·689,279	·685,840	·686,802	·002,150	·002,187	·002,193
·571,117	·568,580	·569,389	·001,666	·001,697	·001,702
·472,553	·470,719	·471,397	·001,285	·001,311	·001,314
·390,437	·389,146	·389,715	·000,987	·001,008	·001,011
·322,111	·321,235	·321,712	·000,753	·000,771	·000,773
·265,332	·264,769	·265,169	·000,572	·000,586	·000,589
·218,213	·217,884	·218,219	·000,432	·000,444	·000,445
·179,164	·179,008	·179,287	·000,324	·000,333	·000,334
·146,850	·146,819	·147,051	·000,241	·000,249	·000,250
·120,150	·120,205	·120,398	·000,178	·000,185	·000,186
·098,123	·098,234	·098,395	·000,131	·000,134	·000,135
·079,979	·080,126	·080,260	·000,095	·000,098	·000,098
·065,060	·065,227	·065,338	·000,068	·000,071	·000,071
·052,814	·052,989	·053,081	·000,050	·000,051	·000,051
·042,779	·042,954	·043,030	·000,034	·000,037	·000,037
·034,572	·034,742	·034,805	·000,024	·000,025	·000,025
·027,874	·028,034	·028,086	·000,016	·000,018	·000,018
·022,418	·022,566	·022,608	·000,011	·000,012	·000,012
·017,983	·018,118	·018,153	·000,007	·000,008	·000,008
·014,386	·014,507	·014,536	·000,005	·000,006	·000,006
·011,476	·011,584	·011,608	·000,003	·000,003	·000,003
·009,128	·009,222	·009,241	·000,002	·000,002	·000,002
·007,237	·007,319	·007,335	·000,001	·000,001	·000,001
·005,719	·005,790	·005,802	·000,000	·000,000	·000,000
·004,504	·004,565	·004,575			

VI. *Series* $F(-100, -100, 1, \frac{1}{2})$; $c = 1$.

The curve which fits the series is

$$y = y_0 e^{-\frac{2R}{3}} e^{\frac{808}{9\sqrt{2}} \arctan \frac{3\sqrt{2}R}{202}} \Big/ \left(R^2 + \frac{20402}{9}\right)^{\frac{808}{9}},$$

where $R = z - 101/3$. $z_{(\text{mode})} = 41{\cdot}835{,}570$.

Significant Terms of Series $\times 10^{-45}$	Corresponding Mid-ordinates of Curve $\times 10^{-45}$	Areas under Curve $\times 10^{-45}$	Significant Terms of Series $\times 10^{-45}$	Corresponding Mid-ordinates of Curve $\times 10^{-45}$	Areas under Curve $\times 10^{-45}$
·000,002	·000,004	·000,004	·165,172	·164,379	·163,910
·000,007	·000,014	·000,015	·138,596	·137,167	·137,060
·000,027	·000,046	·000,048	·107,318	·107,009	·106,997
·000,093	·000,137	·000,142	·076,710	·076,730	·076,956
·000,287	·000,379	·000,393	·050,631	·051,513	·051,719
·000,803	·000,977	·001,008	·030,864	·031,366	·031,656
·002,048	·002,341	·002,401	·017,380	·017,935	·018,145
·004,762	·005,192	·005,300	·009,041	·009,542	·009,690
·010,110	·010,658	·010,828	·004,345	·004,731	·004,825
·019,630	·020,215	·020,451	·001,929	·002,190	·002,243
·034,900	·035,391	·035,669	·000,791	·000,949	·000,976
·056,888	·057,142	·057,404	·000,300	·000,386	·000,399
·085,104	·085,212	·085,183	·000,105	·000,147	·000,153
·116,959	·116,539	·116,503	·000,034	·000,053	·000,055
·147,794	·147,110	·146,830	·000,010	·000,018	·000,019
·171,857	·171,012	·170,664	·000,003	·000,006	·000,006
·184,023	·183,113	·182,497	·000,001	·000,002	·000,002
·181,571	·180,676	·180,092			

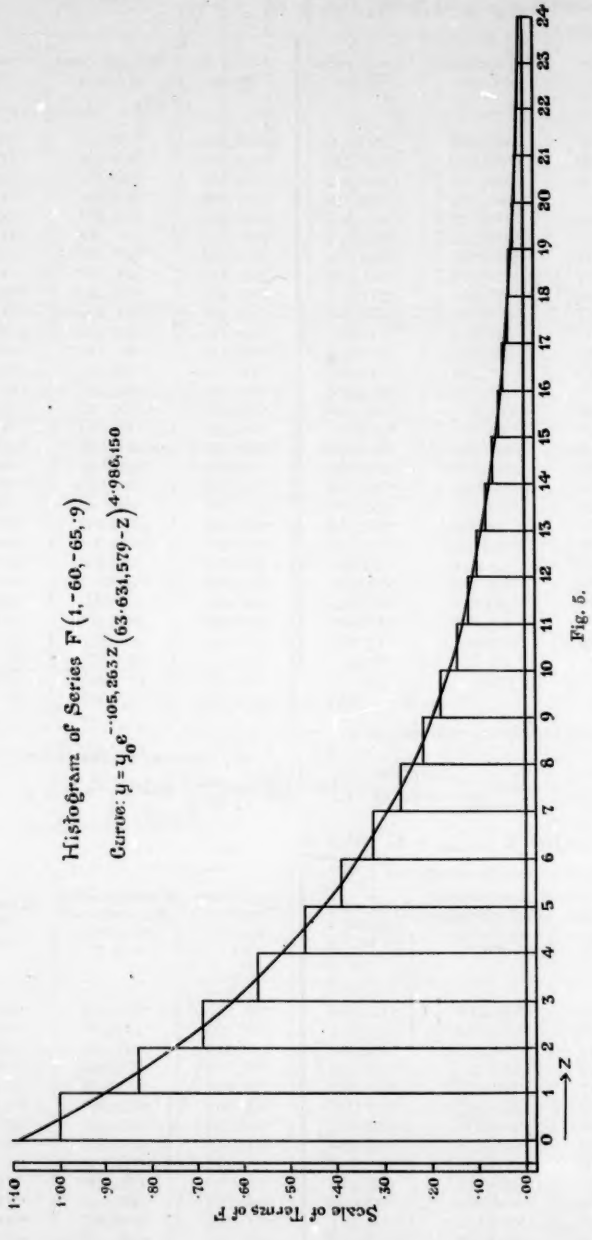

Histogram of Series F(1, -60, -65, ·9)
Curve: $y = y_0 e^{-105{\cdot}263z}(63{\cdot}631{\cdot}579 - z)^{4{\cdot}986{\cdot}150}$
Scale of Terms of F
1·10
1·00
·90
·80
·70
·60
·50
·40
·30
·20
·10
·00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
z

Fig. 5.

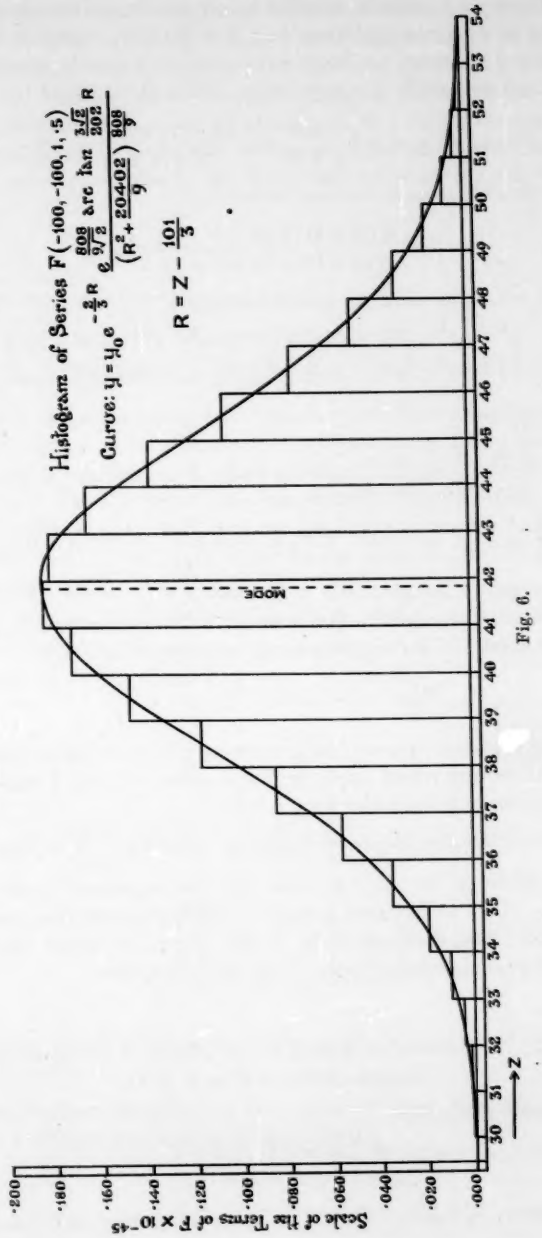

Histogram of Series F(-100, -100, 1, ·5)
Curve: $y = y_0 e^{-\frac{2}{3}R} \frac{e^{\frac{808}{9\sqrt{2}} \text{ arc tan } \frac{3\sqrt{2}}{202} R}}{\left(R^2 + \frac{20402}{9}\right)^{\frac{808}{9}}}$
$R = Z - \frac{101}{3}$
MODE
Scale of the Terms of F × 10⁻⁴³
·200
·180
·160
·140
·120
·100
·080
·060
·040
·020
·000
30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
Z

Fig. 6.

The condition $x \leqslant 1$ must be satisfied by all convergent infinite hypergeometric series; but for finite series (i.e., when α or β or both are negative integers) x may have any value whatsoever. A finite series with $x>1$ may be reversed into one in which the fourth element is less than unity. Thus, if the series be

$$F(-\alpha', \beta, \gamma, x) = 1 - \frac{\alpha'\beta}{1!\gamma} + \frac{\alpha'(\alpha'-1)\beta(\beta+1)}{2!\gamma(\gamma+1)}x^2 - \ldots,$$

where α' is a positive integer, the last term is

$$(-)^{\alpha'}\frac{\beta(\beta+1)\ldots(\beta+\alpha'-1)}{\gamma(\gamma+1)\ldots(\gamma+\alpha'-1)}x^{\alpha'} = \chi.$$

Reversing the series, we have the equivalent one

$$\chi\left[1 - \frac{\alpha'\gamma+\alpha'-1}{1!\beta+\alpha'-1}\frac{1}{x} + \frac{\alpha'(\alpha'-1)(\gamma+\alpha'-1)(\gamma+\alpha'-2)}{2!(\beta+\alpha'-1)(\beta+\alpha'-2)}\frac{1}{x^2} - \ldots\right],$$

i.e.,

$$F(-\alpha', \beta, \gamma, x)$$
$$= (-)^{\alpha'}\frac{\beta(\beta+1)\ldots(\beta+\alpha'-1)}{\gamma(\gamma+1)\ldots(\gamma+\alpha'-1)}F\left\{-\alpha', -(\gamma+\alpha'-1), -(\beta+\alpha'-1), \frac{1}{x}\right\},$$

i.e. $$F(\alpha, \beta, \gamma, x) = \text{const.} \times F\left(\alpha, -\gamma+\alpha+1, -\beta+\alpha+1, \frac{1}{x}\right),$$

where α is a negative integer. We notice that $\epsilon = \gamma-\alpha-\beta-1$ is the same for both series. Replacing z of the fundamental differential equation ((2), p. 61) by $-z-(\alpha-1)c$ results in the corresponding equation for the series

$$F\left(\alpha, -\gamma+\alpha+1, -\beta+\alpha+1, \frac{1}{x}\right).$$

Hence, nothing is gained by considering separately finite series in which x is greater than unity. All results which apply for finite series with $x<1$ apply also, with a slight modification, to finite series with $x>1$.

In the above examples the curve found by equating $\frac{1}{y}\frac{dy}{dz}$ to $\Delta y_r/\frac{1}{2}c(y_{r+1}+y_r)$ of the series, where y_r is the rth term, fits the significant part of the series tolerably well. There is, however, a slight deficiency about the mode and excess towards the tails. The goodness of fit of this system of curves improves rapidly when the number of significant terms of the series increases.

§ 2. *An Alternative Method for the Fitting of Generalised Pearson Curves to* $F(\alpha, \beta, \gamma, 1)$.

Consider firstly the series

$$1 + \frac{\alpha\beta}{1!\gamma} + \frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\gamma(\gamma+1)} + \ldots \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(i)},$$

in which x is unity. When its standard deviation is not small, a Pearson curve $P(z)$

may be fitted to it with sufficient accuracy for most statistical purposes*. Let, then, $P_1(z)$ be the Pearson curve having the same first four moments as (i). Let z_1 be the value of z which corresponds to the mid-ordinate of the first block of the histogram of series (i). The abscissa of the mid-vertical of the rth block is, therefore, $P_1(z_1+r)$, the grouping unit c being taken as unity. $P_1(z+r)$ also gives quite closely the numerical value of the rth term of the series (i). Hence, the expression $x^{r-1}P_1(z_1+r)$ gives quite closely the numerical value of the rth term of the general series

$$1+\frac{\alpha\beta}{1!\gamma}x+\frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\gamma(\gamma+1)}x^2+\ldots \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(ii)}.$$

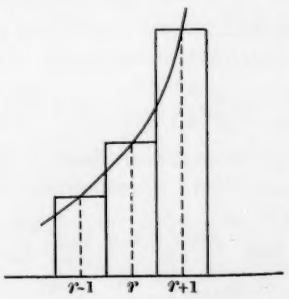
Accordingly, the generalised Pearson curve

$$x^{z-z_1-1}P_1(z)=\text{const.}\,x^z P_1(z)$$

will give quite closely the *ordinates* of the general series.

We know that the areas under the Pearson curve $P(z)$ accord more closely to the terms of the series (i) than the ordinates*, but owing to the presence of the factor x^z, or what is equivalent, $e^{\rho z}$ where $\rho=\log_e x$, this property no longer holds for the generalised Pearson curve. We can, however, apply a small correction so that the *areas* under $e^{\rho z}P_1(z)$ will accord to the terms of the series (ii) at least as closely as do the areas under $P_1(z)$ accord to the terms of (i). This correction is obtained as follows:

Let Y_r denote the ordinate of $x^z P_1(z)$ which corresponds to the mid-point of the $(r+1)$st block of the histogram of the general series, and let y_r be the corresponding ordinate of $P_1(z)$. Then $Y_r=x^r y_r$.



Let N_r denote the $(r+1)$st term of the series (ii) and n_r the $(r+1)$st term of (i). Then, $N_r=x^r n_r$.

When the standard deviation of the series is large, the following formula expresses with sufficient accuracy, n_r in terms of y_r:

$$n_r=\tfrac{1}{24}(22y_r+y_{r-1}+y_{r+1}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(iii)}.$$

* See O. L. Davies, *Biometrika*, Vol. xxv. pp. 295—322.

The area under the generalised Pearson curve corresponding to the $(r+1)$st term of the series (ii) is

$$\tfrac{1}{24}(22Y_r + Y_{r-1} + Y_{r+1}) \quad \text{.....(iv).}$$

Now
$$N_r = x^r n_r = \frac{x^r}{24}(22y_r + y_{r-1} + y_{r+1})$$
$$= \tfrac{1}{24}(22Y_r + Y_{r-1}x + Y_{r+1}/x) \quad \text{.....(v).}$$

Hence, the correction to be applied to the $(r+1)$st sectional area (iv) is

$$c_r = -\tfrac{1}{24}[Y_{r-1}(1-x) + Y_{r+1}(1-1/x)]$$
$$= \frac{-(1-x)}{24x}[xY_{r-1} - Y_{r+1}] \quad \text{.....(vi).}$$

If, therefore, Na_t denotes the whole area of the curve $P_1(z)x^z$ up to and including the $(t+1)$st term $\left(= \int_0^{t+1} Y_z dz\right)$, the correction for Na_t is

$$\frac{1-x}{24x}\sum_{r=0}^{t}(Y_{r+1} - xY_{r-1})$$
$$= \frac{1-x}{24x}\left[\sum_{r=0}^{t} Y_r(1-x) + (Y_{t+1} + xY_t) - (Y_0 + xY_{-1})\right] \quad \text{.....(vi}b\text{).}$$

If there is high contact at the start of the curve, y_{-1} and y_0 are very nearly zero and the correction simplifies to

$$\frac{1-x}{24x}\left[\sum_{r=0}^{t} Y_r(1-x) + (xY_t + Y_{t+1})\right].$$

Now $\sum_{r=0}^{t} Y_r = Na_t + E$, where E is a small correction found by applying the Euler-Maclaurin theorem.

The correction to be applied to Na_t so that it may give a closer approximation to the sum of the first $(t+1)$ terms of the series (ii) is, therefore,

$$C_r = \left(\frac{1-x}{24x}\right)[(Na_t + E)(1-x) + (xY_t + Y_{t+1})] \quad \text{.....(vii).}$$

E is of the second degree of smallness compared with C_r, and when x is not small it may be neglected altogether. C_r may then be written in the form

$$\frac{(1-x)^2}{24x}Na_t + (xY_t + Y_{t+1})\frac{1-x}{24x} \quad \text{.....(viii),}$$

which may be calculated quite readily.

If there be high contact at both ends, the correction for the whole area of the curve reduces to $\frac{(1-x)^2}{24x}N$. A close approximation to the sum of the whole series is, therefore,

$$N\left[1 + \frac{(1-x)^2}{24x}\right],$$

where N is the total area of the curve $Y = x^z P_1(z)$, provided x is not small. When x approaches unity, the correction tends to zero.

If the simple formula (iii) does not give sufficient accuracy, we may apply the higher quadrature formula

$$n_r = \tfrac{1}{5760}\,[5178y_r + 308\,(y_{r+1} + y_{r-1}) - 17\,(y_{r+2} + y_{r-2})].$$

The correction to be applied to the $(r+1)$st sectional area is

$$\frac{(1-x)}{5760x}\left[308\,(Y_{r+1} - x\,Y_{r-1}) - 17\,\frac{1+x}{x}(Y_{r+2} - x^2 Y_{r-2})\right].$$

Accordingly, the correction to be applied to the area Na_t is

$$\frac{(1-x)^2}{5760x}\,Na_t\left[308 - 17\,\frac{(1+x)^2}{x}\right]$$
$$+\frac{(1-x)}{5760x}\left[308\,(Y_{t+1} + x\,Y_t) - 17\left(\frac{1+x}{x}\right)\{(Y_{t+1} + Y_{t+2}) - x^2\,(Y_{t-1} + Y_t)\}\right]$$
$$-\frac{(1-x)}{5760x}\left[308\,(Y_0 + x\,Y_{-1}) + 17\left(\frac{1+x}{x}\right)\{(Y_{-1} + Y_{-2})\,x^2 + (Y_1 + Y_0)\}\right].$$

When there is high contact at the start of the curve, the last term approximates to zero. Further, when there is high contact at both ends of the curve, the correction for the whole area is

$$\frac{(1-x)^2}{5760x}\,N\left\{308 - 17\,\frac{(1+x)^2}{x}\right\},$$

and accordingly, an approximation to the sum of the general series (ii) is

$$N\left[1 + \frac{(1-x)^2}{5760x}\left\{308 - 17\,\frac{(1+x)^2}{x}\right\}\right],$$

where N is the total area of the curve $x^z P_1(z)$.

The following results will give an indication of the magnitude of the correction C for the whole area when there is high contact at both ends.

$x = \cdot5$, $\quad C = (\frac{4630}{2304})\,\%\cong 2\,\%$,

$x = \cdot75$, $\quad C \cong \cdot34\,\%$,

$x = \cdot9$, $\quad C \cong \cdot046\,\%$.

In the following examples, the areas under the curves obtained by both methods are compared with the terms of the series.

The curve obtained by applying the first method, i.e. from the equation

$$\frac{1}{y}\frac{dy}{dz} = \frac{\Delta y_r}{\frac{c}{2}\,(y_r + y_{r+1})}$$

(§ 1), will be referred to as curve A, while the one obtained by applying the above method will be referred to as curve B.

Examples.

I. *Series* $F(30, 30, 111, \cdot75)$; $c = 1$.

Equation of curve A is

$$y = y_0 e^{-\frac{3}{7}z}\,(z + 4{\cdot}322{,}090)^{13{\cdot}203{,}457}\,(z + 83{\cdot}392{,}196)^{-84{\cdot}142{,}232}.$$

Equation of curve B is

$$y = y_0' e^{-\cdot 287,682z} (z + 4\cdot 408,767)^{13\cdot 408,468} (z + 90\cdot 876,152)^{-69\cdot 792,127}.$$

The origin of both curves is at the start of the histogram of the series.

Terms of Series	Areas under A	Corrected Areas under B	Terms of Series	Areas under A	Corrected Areas under B
1·000	1·866	2·143	43·152	43·638	43·174
6·081	7·539	7·592	31·030	31·496	31·036
19·567	21·166	21·147	21·877	22·299	21·875
44·328	45·503	45·344	15·152	15·517	15·146
79·397	79·826	79·539	10·327	10·632	10·321
119·717	119·421	119·114	6·937	7·183	6·932
158·031	157·252	157·067	4·599	4·937	4·596
187·553	186·573	186·615	3·013	3·160	3·010
203·994	203·042	203·317	1·952	2·062	1·951
206·280	205·490	205·922	1·252	1·334	1·252
196·095	195·535	196·001	·796	·896	·796
176·795	176·484	176·865	·502	·544	·502
152·250	152·178	152·392	·314	·343	·314
125·971	126·107	126·115	·195	·215	·195
100·628	100·928	100·745	·121	·134	·121
77·926	78·342	78·005	...	...	...
58·705	59·176	58·752			

II. *Series* $F(-30, -50, 100, \frac{1}{2})$; $c = 1$.

Equation of curve A is

$$y = y_0 e^{-\frac{3}{2}z} \left\{\left(z + \frac{58}{3}\right)^2 + \frac{1379}{9}\right\}^{-\frac{2172}{27}} e^{\frac{121928}{9\sqrt{(1379)}} \arctan \frac{3}{\sqrt{(1379)}}\left(z + \frac{58}{3}\right)}.$$

Equation of curve B is

$$y = y_0' e^{-\cdot 693,147z} (z + 6\cdot 051,499)^{27\cdot 737,858} (36\cdot 453,380 - z)^{52\cdot 069,623},$$

origin of both curves being at the start of the histogram.

Terms of Series	Areas under A	Corrected Areas under B	Terms of Series	Areas under A	Corrected Areas under B
1·000	2·124	1·506	17·899	18·921	17·909
7·500	9·324	8·157	7·069	8·043	7·072
26·380	27·538	26·523	2·337	2·976	2·341
57·932	57·201	57·502	·650	·974	·653
89·219	86·861	88·893	·153	·285	·154
102·602	99·994	102·742	·030	·076	·031
91·609	89·536	91·946	·005	·018	·005
65·188	64·423	65·408	...	...	...
37·658	38·098	37·728			

III. *Series* $F(-30, 60, -81, \frac{1}{3})$; $c = 1$.

Equation of curve A is

$$y = y_0 e^{-z} (z + 7\cdot 388,310)^{28\cdot 599,628} (61\cdot 888,310 - z)^{53\cdot 900,372}.$$

Equation of curve B is

$$y = y_0' e^{-1{\cdot}098,612z} (z + 7{\cdot}531,867)^{27{\cdot}629,947} (37{\cdot}648,329 - z)^{24{\cdot}035,187},$$

origin of both curves being at the start of the histogram.

Terms of Series	Areas under A	Corrected Areas under B	Terms of Series	Areas under A	Corrected Areas under B
1·000	2·318	1·937	92·107	92·405	92·144
7·407	11·512	10·646	55·036	56·115	55·074
27·299	29·901	30·477	29·462	30·803	29·478
66·655	67·817	70·564	14·189	15·388	14·191
121·133	120·174	123·146	6·165	7·040	6·164
174·515	172·412	174·767	2·421	2·963	2·420
207·300	203·876	206·642	·860	1·152	·859
208·485	205·333	207·814	·276	·416	·276
180·898	178·696	180·567	·080	·138	·080
137·303	136·405	137·235	...	...	...

IV. *Series* $F(1, -60, -65, \cdot 9)$; $c = 1$.

Equation of curve A is

$$y = y_0 e^{-\cdot 105,263 2z} (63{\cdot}631,579 - z)^{4{\cdot}986,150}.$$

Equation of curve B is

$$y = y_0' e^{-\cdot 105,360 5z} (63{\cdot}349,729 - z)^{4{\cdot}988,824},$$

origin of both curves being at the start of the histogram of the series.

Terms of Series	Areas under A	Corrected Areas under B	Terms of Series	Areas under A	Corrected Areas under B
1·000,000	·995,309	·999,980	·003,534	·003,595	·003,534
·830,769	·827,323	·830,630	·002,762	·002,813	·002,762
·689,279	·686,802	·689,193	·002,150	·002,193	·002,150
·571,117	·579,389	·571,058	·001,666	·001,702	·001,666
·472,553	·471,397	·472,519	·001,285	·001,314	·001,285
·390,437	·389,715	·390,421	·000,987	·001,011	·000,987
·322,111	·321,712	·322,105	·000,753	·000,773	·000,753
·265,332	·265,169	·265,334	·000,572	·000,589	·000,572
·218,213	·218,219	·218,221	·000,432	·000,445	·000,432
·179,164	·179,287	·179,174	·000,324	·000,334	·000,324
·146,850	·147,051	·146,863	·000,241	·000,250	·000,241
·120,150	·120,398	·120,162	·000,178	·000,186	·000,178
·098,123	·098,395	·098,136	·000,131	·000,135	·000,131
·079,979	·080,260	·079,990	·000,095	·000,098	·000,095
·065,060	·065,338	·065,071	·000,068	·000,071	·000,068
·052,814	·053,081	·052,823	·000,049	·000,051	·000,049
·042,779	·043,030	·042,786	·000,034	·000,037	·000,034
·034,572	·034,805	·034,579	·000,024	·000,025	·000,024
·027,874	·028,086	·027,879	·000,016	·000,018	·000,016
·022,418	·022,608	·022,422	·000,011	·000,012	·000,011
·017,983	·018,153	·017,987	·000,007	·000,008	·000,007
·014,386	·014,536	·014,389	·000,005	·000,006	·000,005
·011,476	·011,608	·011,479	·000,003	·000,003	·000,003
·009,128	·009,241	·009,129	·000,002	·000,002	·000,002
·007,237	·007,335	·007,238	·000,001	·000,001	·000,001
·005,719	·005,802	·005,720	·000,000	·000,000	·000,000
·004,504	·004,575	·004,504	...	...	...

There is very little difference between the fit of the two curves at points up to the mode, but after the mode and especially towards the tail, the curve B is a distinct improvement on curve A. In the last example, where the curve is J-shaped, the former method gives far better results than that developed in § 1. An added advantage is that the fairly troublesome type G_{IV} does not arise. It must be remembered, however, that this method is applicable only when the series with x replaced by unity is convergent.

§ 3. *Evaluation of the Incomplete Integral* $\int_{z_1}^{z_2} e^{-\rho z} P(z)\, dz$.

In the previous paragraphs we have shown that when the number of significant terms is fairly large, a general hypergeometric series $F(\alpha, \beta, \gamma, x)$ may be fitted quite closely by a curve of the form

$$y = y_0 e^{-\rho z} P(z) \quad \text{.....(1),}$$

where $P(z)$ is a Pearson type curve. Hence, a number of terms of this series can be represented by an integral of y taken between the proper limits. When this curve is not very skew, we may often evaluate the integral by using the method outlined by Pearson in *Biometrika*, vol. I. p. 390 *et seq.* This consists in expanding (1) about its mode into a series of the form

$$y = y_m e^{\frac{1}{2}\left(\frac{d^2u}{dz^2}\right)_m z^2} (1 + b_3 z^3/3! + b_4 z^4/4! + \ldots),$$

where $u = \log y$, and $z = m$, the position of the mode of the curve y. The coefficients are expressed by Pearson in the form

$$b_3 = a_3, \qquad b_6 = (a_6 + 10a_3^2),$$
$$b_4 = a_4, \qquad b_7 = (a_7 + 35a_3a_4),$$
$$b_5 = a_5, \qquad b_8 = (a_8 + 56a_3a_5 + 35a_4^2), \text{ etc.}$$

where $$a_i = \left(\frac{d^iu}{dz^i}\right)_{z=m} \qquad (i = 1, 2, \ldots).$$

These coefficients become increasingly more complicated.

When finding the partial integrals of the curve, i.e., integrals from the start of the curve up to a point z, we may safely take the lower limit to be $-\infty$. This enables us to express the partial areas of (1) in terms of incomplete normal moment functions. These have been tabulated up to the twelfth. (See *Tables for Statisticians and Biometricians*, Pts. I and II.)

If l_1 and l_2 are the limits to the range of the curve y_1 the complete integral is

$$N = \int_{l_1}^{l_2} y\, dz \cong \int_{-\infty}^{+\infty} y\, dz$$
$$= \frac{y_m \sqrt{(2\pi)}}{\sqrt{(-a_2)}} \left[1 + \frac{b_4}{8a_2^2} - \frac{b_6}{48a_2^3} + \ldots\right] \quad \text{.....(2).}$$

This method will be referred to as Method A.

Example.

$$y = y_0 e^{-z} z^{30} (40 - z)^{60}.$$

The total area of the curve found by quadrature is $79345{\cdot}112 y_0 10^{110}$.

It is convenient to take $y_0 = 10^{-110}$.

Now $$u = \log y \quad \text{and} \quad \frac{du}{dz} = -1 + \frac{30}{z} - \frac{60}{40 - z}.$$

The mode is, therefore, at the point $z = 10$. The constants

$$a_q = \left(\frac{d^q u}{dz^q}\right)_{10} = (q-1)! \left\{(-)^q \frac{30}{10^q} - \frac{60}{30^q}\right\} \qquad (q = 2, 3, \ldots)$$

are readily found. Substitute their values in (2) and we have

$$\begin{aligned} N &= 79668{\cdot}590\,(1 - {\cdot}017{,}14876 + {\cdot}011{,}51807 \\ &\qquad + {\cdot}004{,}69824 - {\cdot}003{,}31286 + {\cdot}000{,}90737 - \ldots) \\ &= 79402{\cdot}661. \end{aligned}$$

This is in excess of the true value by approximately $\cdot 072\,{}^0/_0$, which is sufficiently close for most statistical purposes.

Method A has been applied* with varying success to the evaluation of the incomplete Beta function. The results obtained there hold with a slight modification for integrals of the curve $e^{-pz} P(z)$ because this curve and $P(z)$ behave in very much the same manner. It has been shown that the curve obtained by Method A fits a Pearson type curve very well at points lying within $\pm 1{\cdot}5\sigma$ of the mode, but may deviate appreciably at points further removed from the mode. This is also the case with the more general curves.

In the next three paragraphs we shall develop other and more accurate methods of curve fitting.

§ 4. *The Fitting of Pearson Curves to Generalised Pearson Curves by Equating the Corresponding Logarithmic Differentials about the Mode.*

Method A may be generalised somewhat by expanding

$$y = y_0 e^{-pz} P(z)$$

around the mode into a series of the form

$$y = y_0 u(z) \{1 + b_1 z + b_2 z^2 + \ldots\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(i)},$$

where $u(z)$ is a Pearson curve of the same type as $P(z)$.

When $P(z)$ is a Type I curve, y may be expressed in the form

$$y = y_0 e^{-pz} z^r (a - z)^s \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(ii)}.$$

Let (ii) be represented around its mode "m" by the curve

$$y = y_0' (z + m)^p (a - m - z)^q \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(iii)},$$

where p and q are at our disposal.

* *Biometrika*, Vol. VI. (1908), p. 68; *Introduction to Incomplete* Γ-*Function Tables* (1922), pp. xviii—xix; *Tables for Statisticians*, Part II. (1931), pp. ccxxxi—ccxxxii; *Tables of the Incomplete* B-*Function* (1934), p. vi; Soper, *Tracts for Computers*, VII. (1921), pp. 43—44; Wishart, *Biometrika*, Vol. XIX. (1927), p. 438.

Now
$$y = y_0'' e^{-\rho z}(z+m)^r (a-m-z)^s$$
$$= y_0''' (m+z)^p (a-m-z)^q V,$$
where
$$V = e^{-\rho z}(m+z)^{r-p}(a-m-z)^{s-q}$$
and m the distance of the mode of (ii) from the start of the curve.

Let
$$v = \log V,$$
then
$$(dV/dz) = V(dv/dz),$$
$$(d^2V/dz^2) = V(d^2v/dz^2) + V(dv/dz)^2, \text{ etc.}$$

Choose p and q so that
$$(dv/dz)_m = 0 \quad \text{.....(iv } a\text{)},$$
$$(d^2v/dz^2)_m = 0 \quad \text{.....(iv } b\text{)}.$$

The two curves (ii) and (iii) have the same range. Condition (iv a) implies that they also have the same mode and condition, (iv b) indicates that the two curves coincide to the second order around the mode.

By Maclaurin's theorem
$$y = y_0^{(\text{iv})}(m+z)^p (a-m-z)^q \times \left\{1 + \frac{1}{6}\left(\frac{d^3v}{dz^3}\right)_m z^3 + \frac{1}{24}\left(\frac{d^4v}{dz^4}\right)_m z^4 + \frac{1}{120}\left(\frac{d^5v}{dz^5}\right)_m z^5 + \ldots\right\}.$$

The coefficients of the successive powers of z have the same form as those for the expansion obtained by Method A. Here, however,
$$a_i = (d^i \log V/dz^i)_m; \quad V = e^{-\rho z}(m+z)^{r-p}(a-m-z)^{s-q}.$$

The partial integral $y_0 \int_0^z e^{-\rho z} z^r (a-z)^s dz$ may now be expressed in terms of the incomplete moments of Type I curves about the mode which may, in turn, be expressed in terms of incomplete Beta functions. These have been tabulated within a certain range, and, for functions lying outside this range, auxiliary tables have been constructed to aid in their evaluation.

The tth moment of $(m+z)^p (a-m-z)^q$ about its mode is
$$\int_{-m}^{a-m} (m+z)^p (a-m-z)^q z^t dz$$
$$= (-)^t \int_0^a z^p (a-z)^q (m-z)^t dz$$
$$= (-)^t \sum_{i=0}^{\infty} (-)^i m^{t-i} \int_0^a z^{p+i}(a-z)^q \{t!/i!(t-i)!\} dz$$
$$= (-)^t a^{p+q+1} \frac{\Gamma(p+1)\Gamma(q+1)}{\Gamma(p+q+2)} \left[m^t - t m^{t-1} \frac{(p+1)a}{(p+q+2)} + \frac{t(t-1)}{2!} m^{t-2} \frac{(p+1)(p+2)}{(p+q+2)(p+q+3)} a^2 - \ldots + (-)^t \frac{(p+1)(p+2)\ldots(p+t)a^t}{(p+q+2)(p+q+3)\ldots(p+q+t+1)}\right]$$
$$\equiv (-)^t a^{p+q+1} \frac{\Gamma(p+1)\Gamma(q+1)}{\Gamma(p+q+2)} M_t.$$

Hence the total area N is equal to

$$y_m m^{-p}(a-m)^{-q}a^{p+q+1}\frac{\Gamma(p+1)\Gamma(q+1)}{\Gamma(p+q+2)}[1-b_3M_3+b_4M_4-b_5M_5+\ldots] \quad \text{(v)}.$$

When evaluating the incomplete integral of y, M_t has to be replaced by the corresponding series of incomplete Beta function ratios.

Examples.

I. $$y = y_0 e^{-z} z^{30}(40-z)^{60};$$
$$= y_0 z^p (40-z)^q V; \quad y_0 = 10^{-110},$$

where $$V = e^{-z} z^{30-p}(40-z)^{60-q}.$$

Choose those values of p and q which satisfy

$$0 = (d \log V/dz)_m = -1 + \frac{(30-p)}{10} - \frac{(60-q)}{30},$$

$$0 = (d^2 \log V/dz^2)_m = -\frac{(30-p)}{100} - \frac{60-q}{900};$$

i.e. $$p = 27{\cdot}5, \quad q = 82{\cdot}5.$$

The constants $a_t = (d^t \log V/dz^t)_m$ may be readily calculated from

$$a_t = (t-1)!\left[(-)^{t+1}\frac{(30-p)}{10^t} - \frac{(60-q)}{30^t}\right]$$
$$= \frac{(t-1)!\,2{\cdot}5}{10^t}[(-)^{t+1} + (\tfrac{1}{3})^{t-2}].$$

Substitute these values in (v) and we have the following series for the total area:

$$N = 79210\,(1 + {\cdot}002{,}63380 - {\cdot}001{,}27338 + {\cdot}000{,}22961$$
$$+ {\cdot}000{,}06981 - {\cdot}000{,}02788 + {\cdot}000{,}04546 - \ldots)$$
$$= 79342{\cdot}867.$$

This involves an error of ·0028 %. When we include seven terms in the expansion of y, the resulting curve fits y very well at points lying within the range $\pm 2{\cdot}5\sigma$ of the mode, but may deviate somewhat at points lying outside this range.

II. $y = y_0 e^{-\frac{1}{2}z} z^{100}(150-z)^{150}$, $y_0 = 10^{-320}e^{25}$ (mode at $z = 50$).

We have $y = y_0'(z+50)^p(100-z)^q(1 + a_3 z^3/3! + a_4 z^4/4! \ldots)$,

where a_i is the ith logarithmic derivative of V at $z = 0$, V being the expression

$$e^{-\frac{1}{2}z}(z+50)^{100-p}(100-z)^{150-q}.$$

The values of p and q found by equating to zero the first two derivatives of $\log V$ at $z = 0$, are

$$p = 275/3, \quad q = 2p = 550/3.$$

The series giving the total area N is

$$N = c\,(1 + {\cdot}0005{,}39144 - {\cdot}0002{,}44637 + {\cdot}0000{,}12131$$
$$+ {\cdot}0000{,}44450 - {\cdot}0000{,}02193 - \ldots)$$
$$= c\,(1{\cdot}000{,}353282) \text{ and } c = 8{\cdot}409{,}952.$$

Therefore $$N = 8{\cdot}412923.$$

The area of the curve found by quadrature is 8·412901. Hence N differs from the true value by less than ·0003 °/₀. Had we taken $N=c$, i.e. neglected all terms except the first in the expansion, the error would only have been ·03 °/₀. This in itself is sufficiently accurate for most statistical purposes. If we include six terms in the expansion of y, the resulting curve fits y remarkably well at points within $\pm 3\sigma$ of the mode.

The series obtained by the application of Method A is

$$N' = c'(1 - \cdot 004{,}33884 + \cdot 002{,}01703 \ldots).$$

Its convergence is not nearly as rapid as the series obtained by the method of this section.

Although the latter method gives far greater accuracy than Pearson's, it has the serious drawback in being extremely laborious in application, especially when attempting to find partial areas. For this reason we will develop a method of fitting a Pearson type curve which depends on the equating of the first four logarithmic derivatives at the mode. It is hoped that the resulting fit will be sufficiently good for statistical purposes. An incomplete integral of the original curve would then, with the exception of certain cases, be expressible in terms of a single incomplete Beta function.

§ 5. *A systematic Fitting of Pearson Curves to certain Frequency Distributions.*

Let $f(z)$ be a bell-shaped distribution and let $a_1, a_2, a_3, \ldots$ be its successive logarithmic differentials about its mode.

Let y be a Pearson type curve referred to its mode; its differential equation is then

$$\frac{1}{y}\frac{dy}{dz} = \frac{d\log y}{dz} = \frac{z}{b_0 + b_1 z + b_2 z^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(i)}.$$

Differentiate both sides of (i) successively with respect to z, we have

$$\frac{d^2 \log y}{dz^2} = \frac{1}{b_0 + b_1 z + b_2 z^2} - \frac{z\,(b_1 + 2b_2 z)}{(b_0 + b_1 z + b_2 z^2)^2},$$

$$\frac{d^3 \log y}{dz^3} = -\frac{2b_1 + 6b_2 z}{(b_0 + b_1 z + b_2 z^2)^2} + \frac{2z\,(b_1 + 2b_2 z)^2}{(b_0 + b_1 z + b_2 z^2)^3},$$

$$\frac{d^4 \log y}{dz^4} = -\frac{6b_2}{(b_0 + b_1 z + b_2 z^2)^2} + \frac{2\,(3b_1^2 + 18 b_1 b_2 z + 24 b_2^2 z^2)}{(b_0 + b_1 z + b_2 z^2)^3} - \frac{6z\,(b_1 + 2b_2 z)^3}{(b_0 + b_1 z + b_2 z^2)^4},$$

etc.

Equate now the corresponding logarithmic differentials of $y(z)$ and $f(z)$ at their respective modes. We have the following set of equations to solve for the unknown b's:

$$\begin{aligned} 1/b_0 &= a_2, \\ -2b_1/b_0^2 &= a_3, \\ 6\,(b_1^2 - b_0 b_2)/b_0^3 &= a_4. \end{aligned}$$

Hence
$$b_0 = 1/a_2,$$
$$b_1 = -a_3/2a_2^2,$$
$$b_2 = (3a_3^2 - 2a_4a_2)/12a_2^3,$$

and
$$\frac{1}{y}\frac{dy}{dz} = \frac{z}{\frac{1}{a_2} - \frac{a_3 z}{2a_2^2} + \frac{1}{12a_2^3}(3a_3^2 - 2a_4a_2)\,z^2}.$$

The discriminant of the quadratic in the denominator is

$$D = \frac{a_3^2}{4a_2^4} - \frac{1}{3a_2^4}(3a_3^2 - 2a_4a_2) = \frac{1}{12a_2^4}\{8a_4a_2 - 9a_3^2\}.$$

Hence the condition for a Type IV curve is ($D < 0$)

$$9a_3^2 > 8a_4a_2.$$

Condition for a Type V curve is ($D = 0$)

$$9a_3^2 = 8a_4a_2.$$

Condition for a Type III curve is ($b_2 = 0$)

$$3a_3^2 = 2a_4a_2.$$

Condition for curves lying above the Type IV line is

$$9a_3^2 < 8a_4a_2.$$

For a Type I curve the roots of the quadratic must be real and of opposite sign, i.e.

$$9a_3^2 < 8a_4a_2,$$

and
$$3a_3^2 < 2a_4a_2.$$

The first condition is here implied in the second.

For a Type VI curve the roots of the quadratic must be real and of the same sign, i.e.

$$9a_3^2 < 8a_4a_2,$$
$$3a_3^2 > 2a_4a_2.$$

The curve is symmetrical when a_3 is zero. The differential equation then reduces to

$$\frac{1}{y}\frac{dy}{dz} = \frac{6a_2^2 z}{6a_2 - a_4 z^2}.$$

a_2 is always negative for bell-shaped curves; hence, the following types arise according to the sign of a_4:

$a_4 < 0$ gives a Type II curve.
$a_4 > 0$ „ „ „ VII „
$a_4 = 0$ „ the normal curve.

Note that the differential equation (i) can never represent a J-shaped curve, when $f(x)$ is bell-shaped.

The above method may be easily generalised to the fitting of higher order Pearson curves, i.e. curves satisfying the differential equation

$$\frac{1}{y}\frac{dy}{dz} = \frac{z}{b_0 + b_1 z + b_2 z^2 + b_3 z^3 + \ldots},$$

by equating the differentials of $\log y$, at $z = 0$, to the corresponding logarithmic differentials of $f(z)$ at its mode. The resulting simultaneous equations may be solved successively for the unknown b's in terms of the known a's.

Table to determine the Type of a Curve.

Main Types	Criterion	Transition Types	Criterion
I	$3a_3^2 < 2a_4a_2$	III	$3a_3^2 = 2a_4a_2$
VI	$8a_4a_2 > 9a_3^2 > 6a_4a_2$	V	$9a_3^2 = 8a_4a_2$
IV	$9a_3^2 > 8a_4a_2$	II	$a_3 = 0,\ a_4 < 0$
—	—	Normal	$a_3 = 0,\ a_4 = 0$
—	—	VII	$a_3 = 0,\ a_4 > 0$

Let $u(z)$ be the Pearson type curve fitted by the above method to $f(z)$. Write $f(z) = u(z)\,V$, where $V = f(z)/u(z)$, both curves being referred to their respective modes. We may expand $\log V$ by Maclaurin's theorem into the following series:

$$\log V = a_0' + a_1' z + a_2' z^2/2! \ldots,$$

where $$a_i' = (d^i \log V/dz^i)_{\text{at mode}}.$$

The first four a's vanish; hence

$$\log V = a_5' z^5/5! + a_6' z^6/6! + \ldots,$$

or
$$V = \exp.\,(a_5' z^5/5! + a_6' z^6/6! + \ldots)$$
$$= 1 + \frac{a_5' z^5}{5!} + \frac{a_6' z^6}{6!} + \ldots + \frac{a_9' z^9}{9!} + \left\{\frac{a_5'^2}{(5!)^2} + \frac{2a_{10}'}{10!}\right\}\frac{z^{10}}{2!} + \ldots.$$

Therefore
$$f(z) = u(z)\,[1 + a_5' z^5/5! + a_6' z^6/6! + \ldots].$$

Examples.

I.
$$y = y_0 e^{-z}(z+10)^{30}(30-z)^{60} \quad \text{(origin at mode)} \ldots\ldots\ldots\ldots\text{(i)}.$$

The first four logarithmic differential coefficients of y at the mode are, respectively,

$$a_1 = 0; \qquad a_2 = -11/30;$$
$$a_3 = 5/90; \qquad a_4 = -166/9000.$$

$3a_3^2$ is less than $2a_4a_2$, hence the curve to be fitted is of Type I. This curve is readily found to be

$$y = y_0'(9{\cdot}828{,}086 + z)^{28{\cdot}210{,}489}(38{\cdot}473{,}920 - z)^{110{\cdot}435{,}344} \ldots\ldots\ldots\text{(ii)},$$
$$\sigma \cong 1{\cdot}65.$$

y_0' is determined by equating the ordinates of (i) and (ii) at their respective modes. The total area of the curve (ii) is ·793,2284, and that of (i), found by mechanical quadrature, is ·793,4511 (y_0 is taken to be $10^{-115} \times e^{10}$)—a difference of ·028 %.

The following table of ordinates is constructed to examine the goodness of fit of the two curves:

z	Ordinates of (i)	Ordinates of (ii)	z	Ordinates of (i)	Ordinates of (ii)
−7	·000,000	·000,000	2	·098,488	·098,493
−6	·000,005	·000,005	3	·045,113	·045,127
−5	·000,277	·000,273	4	·015,926	·015,946
−4	·004,241	·004,227	5	·004,413	·004,429
−3	·026,528	·026,511	6	·000,972	·000,980
−2	·084,595	·084,588	7	·000,171	·000,174
−1	·158,609	·158,608	8	·000,024	·000,025
0	·192,456	·192,456	9	·000,003	·000,003
1	·161,593	·161,593	10	·000,000	·000,000

II. $y = y_0 e^{-\frac{1}{2}z}(z+50)^{100}(100-z)^{150}$, $y_0 = 10^{-320}$ (origin at mode) .....(i).

The first four logarithmic differentials of y at the origin are, respectively,

$$a_1 = 0, \qquad a_2 = -55/1000,$$
$$a_3 = 13/10^4, \qquad a_4 = -105/10^6.$$

Here again $3a_3^2$ is less than $2a_4a_2$, hence the curve having the same first four differentials as y at its mode is of Type I. Its equation is

$$f = y_0'(48{\cdot}73692 + z)^{91{\cdot}740{,}961}(114{\cdot}94062 - z)^{216{\cdot}360{,}891} \text{ .....(ii)},$$
$$\sigma \cong 4{\cdot}25.$$

The total area of the latter curve is 8·41285, which agrees very well with the area (8·412,902) of the curve (i) found by quadrature.

In the following table the ordinates of both curves (i) and (ii) are compared at intervals of 2 for the argument z:

z	Ordinates of (i)	Ordinates of (ii)	z	Ordinates of (i)	Ordinates of (ii)
−22	·000,000	·000,000	4	·514,603	·514,607
−20	·000,001	·000,001	6	·305,561	·305,569
−18	·000,016	·000,016	8	·149,209	·149,225
−16	·000,195	·000,194	10	·060,260	·060,279
−14	·001,607	·001,603	12	·020,221	·020,236
−12	·009,265	·009,254	14	·005,659	·005,668
−10	·038,581	·038,564	16	·001,324	·001,328
− 8	·119,034	·119,018	18	·000,260	·000,261
− 6	·277,991	·277,983	20	·000,043	·000,043
− 4	·500,464	·500,462	22	·000,006	·000,006
− 2	·705,414	·705,414	24	·000,001	·000,001
0	·788,861	·788,861	26	·000,000	·000,000
2	·707,866	·707,869			

III. $$y = y_0 e^{-\frac{1}{2}z}(z+50)^{40}(z+150)^{-70} \quad \text{.....(i)},$$

$$y_0 = 10^{84} \text{ (origin at mode).}$$

The first four logarithmic differentials of (i) at the mode are, respectively,

$$a_1 = 0, \qquad a_2 = -\frac{29}{2250},$$

$$a_3 = \frac{2020}{27\times 50^3}, \qquad a_4 = -\frac{6340}{27\times 50^4}.$$

The Pearson curve which has the same first four logarithmic differentials as (i) is of Type VI, namely,

$$y = y_0'(48{\cdot}456{,}364 + z)^{34{\cdot}589{,}810}(387{\cdot}406{,}732 + z)^{-276{\cdot}544{,}175} \quad \text{.....(ii)},$$

$$\sigma \cong 9{\cdot}05.$$

Area of (i) found by quadrature is 9·53324. Area of (ii) is 9·53364.

The series for the total area, N, deduced by Method A is very uncertain; it is

$$N = 9{\cdot}470{,}962\,(1 - {\cdot}028{,}26991 - {\cdot}031{,}87332 + \ldots).$$

z	Ordinates of (i)	Ordinates of (ii)	z	Ordinates of (i)	Ordinates of (ii)
−32	·000,001	·000,001	20	·059,869	·059,913
−28	·000,051	·000,049	24	·028,607	·028,653
−24	·001,115	·001,100	28	·012,617	·012,657
−20	·010,095	·010,053	32	·005,189	·005,217
−16	·047,650	·047,598	36	·002,007	·002,024
−12	·137,084	·137,056	40	·000,735	·000,745
− 8	·267,859	·267,853	44	·000,257	·000,262
− 4	·384,302	·384,299	48	·000,086	·000,088
0	·428,955	·428,955	52	·000,028	·000,029
4	·389,261	·389,260	56	·000,009	·000,009
8	·297,170	·297,174	60	·000,003	·000,003
12	·196,083	·196,097	64	·000,001	·000,001
16	·114,276	·114,308	68	·000,000	·000,000

The above examples show that the Pearson curve which has the same first four logarithmic differentials about the mode as the curve $y = y_0 e^{-\rho z} P(z)$ (when not very skew) fits it with sufficient accuracy for most statistical purposes at points lying within $\pm 3\sigma$ of the mode. The range of agreement of the two curves is further extended when a few terms of the expansion is included; but these render the finding of partial areas an extremely laborious task. It is desirable, therefore, to develop accurate and convenient methods for fitting integrable curves to the "tail" of these frequency curves.

§ 6. A method similar to the one developed in the beginning of § 3 may generally be applied satisfactorily to determine the area of parts of the curve lying outside the range mode $\pm 2{\cdot}5\sigma$. Thus, let $z \geqslant z_0$ be the tail of the frequency curve

$y = y_0 e^{-pz} P(z)$ and let a_i be the ith logarithmic differential of y at the "stump" $z = z_0$. Then, by Maclaurin,

$$y = e^{a_0 + a_1\zeta + a_2\zeta^2/2! + \ldots}$$
$$= y(z_0)\, e^{a_1\zeta + a_2\zeta^2/2!}\,[1 + a_3\zeta^3/3! + a_4\zeta^4/4! + \ldots],$$

where $\zeta = z - z_0$.

The series in brackets is identical in form with the one deduced by Method A (§ 3).

Evidently,

$$y = y(z_0)\, e^{\frac{1}{2}\frac{z_1^2}{\sigma^2}} e^{-\frac{1}{2\sigma^2}(\zeta + z_1)^2}\,[1 + a_3\zeta^3/3! + a_4\zeta^4/4! + \ldots],$$

where $\sigma = \dfrac{1}{\sqrt{(-a_2)}}$, $z_1 = a_1/a_2$.

Hence, area of tail $z \geqslant z_0$ is

$$N'_{z_0} = y(z_0)\, e^{\frac{1}{2}\frac{z_1^2}{\sigma^2}} \int_0^\infty e^{-\frac{1}{2\sigma^2}(\zeta + z_1)^2}\,[1 + a_3\zeta^3/3! + a_4\zeta^4/4! + \ldots]\, d\zeta.$$

This may be expressed in terms of incomplete normal moment functions. Note that the first approximation is

$$N_{z_0'} = y(z_0)\, e^{\frac{1}{2}\frac{z_1^2}{\sigma^2}} \int_0^\infty e^{-\frac{1}{2\sigma^2}(\zeta + z_1)^2}\, d\zeta$$
$$= y(z_0)\,\sigma \cdot \frac{\frac{1}{2}(1 - \alpha_{z_1'})}{z_{z_1'}},$$

where $z_1' = \dfrac{z_1}{\sigma}$, $\frac{1}{2}(1 - \alpha_{z_1'})$ the area of the tail of a normal curve, and $z_{z_1'}$ the ordinate of a normal curve at the point z_1'. This is the expression obtained by Burton H. Camp* for the determination of the tail of a frequency distribution. Only in rare cases is it necessary to go beyond the 6th term in the above expansion of N.

Examples.

I. *Fit a curve to the tail $z \geqslant 4$ of the curve*

(i) $\qquad y = y_0 e^{-z}(z + 10)^{30}(30 - z)^{60}, \quad y_0 = e^{10}10^{-115}, \quad \sigma \cong 1{\cdot}65$

(origin at the mode).

The first six logarithmic differentials of y at the point $z = 5$ are, respectively,

$$a_1 = -1{\cdot}4, \qquad a_2 = -{\cdot}2293,33333,$$
$$a_3 = \ {\cdot}0100,97777, \qquad a_4 = -{\cdot}0044,77155,$$
$$a_5 = \ {\cdot}0008,0069215, \qquad a_6 = -{\cdot}0003,4554058;$$

and the curve is

(ii) $\qquad y = {\cdot}0044,1287\, e^{-1{\cdot}4\zeta - {\cdot}2293\frac{\zeta^2}{2}}[1 + {\cdot}0016,8296\,\zeta^3$
$\qquad - {\cdot}0001,86548\,\zeta^4 + {\cdot}0000,06672\,\zeta^5$
$\qquad + {\cdot}0000,00936\,\zeta^6 - \ldots],$

where $\zeta = z - 5$.

* *Biometrika*, Vol. xvi. p. 163.

The point $\zeta = -1$ is approximately $2{\cdot}5\sigma$ away from the mode.

The following table compares the ordinates of (i) and (ii) at unit intervals of the argument z:

z	Ordinates of (i)	Ordinates of (ii)
4	·01592,63	·01592,64
5	·00441,29	·00441,29
6	·00097,18	·00097,18
7	·00017,145	·00017,145
8	·00002,433	·00002,435
9	·00000,278	·00000,279
10	·00000,025	·00000,026
11	·00000,002	·00000,002

II. *Fit a curve to the tail $z \geqslant 10$ of the generalised Pearson curve*

(i) $$y = y_0 e^{-\frac{1}{2}z}(z+50)^{100}(100-z)^{150}, \quad y_0 = 10^{-320}, \quad \sigma \cong 4{\cdot}25$$

(origin at the mode).

The first five logarithmic differential coefficients of (i) at $z = 10$ are, respectively,

$$a_1 = -{\cdot}5, \qquad a_2 = -\tfrac{5}{108},$$
$$a_3 = \ {\cdot}00051,44033, \qquad a_4 = -{\cdot}00006,00137,$$
$$a_5 = \ {\cdot}00000,24768.$$

The required curve is

(ii) $$y = {\cdot}0602,600\, e^{a_1\zeta + \frac{a_2}{2}\zeta^2}\left[1 + {\cdot}00068,58106\left(\frac{\zeta}{2}\right)^3 - {\cdot}00004,00091\left(\frac{\zeta}{2}\right)^4 + {\cdot}00000,06605\left(\frac{\zeta}{2}\right)^5 + \ldots\right],$$

where $$\zeta = z - 10.$$

z	Ordinates of (i)	Ordinates of (ii)
10	·0602,600	·0602,600
12	·0202,210	·0202,210
14	·0056,585	·0056,585
16	·0013,241	·0013,240
18	·0002,596	·0002,595
20	·0000,427	·0000,426
22	0000,059	·0000,058
24	·0000,007	·0000,007
26	·0000,001	·0000,001

III. *Fit a curve to the tail of the generalised Pearson curve*

(i) $$y = y_0 e^{-\frac{1}{2}z}(z+50)^{40}(z+150)^{-70} : z \geqslant 20, \quad y_0 = 10^{84}, \quad \sigma \cong 9{\cdot}05$$

(origin at mode).

The first six logarithmic differentials of (i) at $z = 20$ are

$$a_1 = -\cdot 173{,}669{,}468, \qquad a_2 = -\cdot 005{,}741{,}120,$$
$$a_3 = \cdot 000{,}204{,}74032, \qquad a_4 = -\cdot 000{,}009{,}4929675,$$
$$a_5 = \cdot 000{,}000{,}55935840, \qquad a_6 = -\cdot 000{,}000{,}04045132.$$

The required curve is

$$\text{(ii)} \qquad y = \cdot 059{,}869\, e^{a_1\zeta + \frac{a_2}{2}\zeta^2}\left[1 + \cdot 002{,}183{,}897\left(\frac{\zeta}{4}\right)^3 - \cdot 000{,}101{,}258\left(\frac{\zeta}{4}\right)^4 + \cdot 000{,}004{,}773\left(\frac{\zeta}{4}\right)^5 + \cdot 000{,}002{,}154\left(\frac{\zeta}{4}\right)^6 - \ldots\right],$$

where $\zeta = z - 20.$

z	Ordinates of (i)	Ordinates of (ii)
20	·059,869	·059,869
24	·028,607	·028,607
28	·012,617	·012,617
32	·005,189	·005,190
36	·002,007	·002,011
40	·000,735	·000,741
44	·000,257	·000,262
48	·000,086	·000,089
52	·000,028	·000,030
56	·000,008	·000,008

In the above examples, with the possible exception of the last, in which there happens to be an appreciable degree of skewness, the agreement between the two curves is very close indeed. We conclude, therefore, that when the curve

$$G(z) = y_0\, e^{-\rho z}\, P(z)$$

is not very skew, the tail $z \geqslant \xi$, where ξ does not lie within mode $\pm 2{\cdot}5\sigma$, may be fitted very closely by the curve

$$y = G(\xi)\, e^{a_1(z-\xi) + a_2(z-\xi)^2/2}[1 + a_3(z-\xi)^3/3\,! + a_4(z-\xi)^4/4\,! + \ldots],$$

where $a_i = \dfrac{d^i}{dz^i}\{\log G(z)\}$ at $z = \xi$. Hence integrals of $G(z)$ up to any point outside mode $\pm 2{\cdot}5\sigma$ may be expressed in terms of incomplete normal moment functions. The rapidity of convergence of the sequence $\sigma^i a_i/i!$ $(i = 3, 4, 5, \ldots)$ will give a good indication as to how many terms of the expansion y need be retained.

It has already been shown (§ 5) that integrals of $G(z)$ up to points lying within mode $\pm 2{\cdot}5\sigma$ may be expressed very closely by a single incomplete Beta function.

Finally, integrals of very skew and J-shaped curves are best evaluated by an accurate quadrature formula such as Weddle's or Euler-Maclaurin's.

§7. *Methods for Obtaining Approximations to the Sum of a Tail of a Hypergeometric Series.*

When the standard deviation of a hypergeometric series is sufficiently large, the above methods for finding approximations to the sum of a finite number of terms serve very well indeed for regions within $\pm 2{\cdot}5\sigma$ of the mode but do not always give satisfactory results for more remote regions. In problems on probability the tail of a distribution is often of greater importance even than the body itself. For this reason it is desirable to develop more accurate methods for approximating to this tail.

Let, then, the following series of constantly decreasing positive quantities, namely,

$$a_0 + a_1 + a_2 + a_3 + \dots,$$

represent the tail of a hypergeometric series $F(\alpha, \beta, \gamma, x)$.

Draw up the following table of the successive ratios of the terms a_i:

Terms	1st ratios	2nd ratios	3rd ratios	4th ratios	5th ratios
a_0					
	$r_1=\frac{a_1}{a_0}$				
a_1		$p_1=\frac{r_2}{r_1}$			
	$r_2=\frac{a_2}{a_1}$		$q_1=\frac{p_2}{p_1}$		
a_2		$p_2=\frac{r_3}{r_2}$		$u_1=\frac{q_2}{q_1}$	
	$r_3=\frac{a_3}{a_2}$		$q_2=\frac{p_3}{p_2}$		$v_1=\frac{u_2}{u_1}$
a_3		$p_3=\frac{r_4}{r_3}$		$u_2=\frac{q_3}{q_2}$	
	$r_4=\frac{a_4}{a_3}$		$q_3=\frac{p_4}{p_3}$		
a_4		$p_4=\frac{r_5}{r_4}$			
	$r_5=\frac{a_5}{a_4}$				
a_5					
...	...	...	...	...	...

Clearly

$$\frac{a_1}{a_0} = r_1, \quad \frac{a_2}{a_0} = r_1 r_2, \dots, \quad \frac{a_s}{a_0} = r_1 r_2 \dots r_s,$$

$$r_2 = r_1 p_1, \quad r_3 = r_1 p_1 p_2, \dots, \quad r_s = r_1 p_1 p_2 \dots p_{s-1},$$

$$p_2 = p_1 q_1, \quad p_3 = p_1 q_1 q_2, \dots, \quad p_s = p_1 q_1 q_2 \dots q_{s-1}, \qquad \text{etc.}$$

Hence

$$\begin{aligned} a_0 + a_1 + a_2 + a_3 + \dots &= a_0 (1 + r_1 + r_1^2 p_1 + r_1^3 p_1^2 p_2 + \dots r_1^s p_1^{s-1} p_2^{s-2} \dots p_{s-1} + \dots) \\ &= a_0 (1 + r_1 + r_1^2 p_1 + r_1^3 p_1^3 q_1 + r_1^4 p_1^6 q_1^3 q_2 + \dots) \\ &= a_0 (1 + r_1 + r_1^2 p_1 + r_1^3 p_1^3 q_1 + r_1^4 p_1^6 q_1^4 u_1 + \dots) \\ &= a_0 \sum_{s=0} r_1^s \, p_1^{\frac{s(s-1)}{1.2}} \, q_1^{\frac{s(s-1)(s-2)}{1.2.3}} \, u_1^{\frac{s(s-1)(s-2)(s-3)}{1.2.3.4}} \dots \end{aligned}$$

If the first ratios are sensibly constant, $p_1 = q_1 = \ldots = 1$ and the series reduces to the simple geometrical

$$a_0(1 + r_1 + r_1^2 + r_1^3 + \ldots).$$

We may, without fear of appreciable error, take the upper limit to be infinity. Hence *a first approximation* to the sum of the series is

$$S_1 = \frac{a_0}{1 - r_1}.$$

If the second ratios are sensibly constant, i.e. $q_1 = u_1 = v_1 = \ldots = 1$, to the figure of accuracy required in the result, the series reduces to

$$a_0 \sum_{s=0}^{\infty} r_1^s p_1^{\frac{s(s-1)}{1.2}} = a_0(1 + r_1 + r_1^2 p_1 + r_1^3 p_1^3 + r_1^4 p_1^6 + \ldots).$$

The first three terms of the original series suffice to determine the values of the constants a_0, r_1 and p_1. Write $r_1 = r$, $p_1 = 1 - p$. If the tail of the series does not lie too near the mode, p_1 is close to unity, i.e. p is a small quantity. Hence

$$\begin{aligned}
a_0 \sum_{s=0}^{\infty} r^s p_1^{\frac{s(s-1)}{1.2}} &= a_0 \sum_{s=0}^{\infty} r^s (1-p)^{\frac{s(s-1)}{1.2}} \\
&= a_0[1 + r + r^2(1-p) + r^3(1 - 3p + 3p^2 - p^3) \\
&\quad + r^4(1 - 6p + 15p^2 - 20p^3 + 15p^4 - 6p^5 + p^6) \\
&\quad + r^5(1 - 10p + 45p^2 - 120p^3 + 210p^4 - 252p^5 + \ldots) \\
&\quad + \ldots \ldots] \\
&= a_0[1 + r + r^2 + r^3 + \ldots \\
&\quad - pr^2(1 + 3r + 6r^2 + 10r^3 + 15r^4 + \ldots) \\
&\quad + 3p^2r^3(1 + 5r + 15r^2 + 35r^3 + \ldots) \\
&\quad - p^3r^3(1 + 20r + 120r^2 + 455r^3 + \ldots) \\
&\quad + 15p^4r^4(1 + 14r + 91r^2 + \ldots) \\
&\quad + \ldots \ldots].
\end{aligned}$$

The above series may all be summed quite readily. Thus

$$\begin{aligned}
1 + r + r^2 + r^3 + \ldots &= \frac{1}{1-r}, \\
1 + 3r + 6r^2 + 10r^3 + \ldots &= \frac{1}{(1-r)^3}, \\
1 + 5r + 15r^2 + 35r^3 + \ldots &= \frac{1}{(1-r)^5}, \\
1 + 20r + 120r^2 + 455r^3 + \ldots &= \frac{1 + 13r + r^2}{(1-r)^7}, \\
\ldots \quad \ldots \quad &= \quad \ldots.
\end{aligned}$$

Hence *a second approximation* to the sum of the tail is

$$S_2 = \frac{a_0}{1-r}\left[1 - p\left(\frac{r}{1-r}\right)^2 + 3p^2\frac{r^3}{(1-r)^4} - \frac{r^3p^3(1 + 13r + r^2)}{(1-r)^6}\right].$$

The first three terms are generally sufficient for computing purposes. The next approximation will involve an additional constant q_1.

$$S_3 = a_0 (1 + r + r^2 p_1 + r^3 p_1^3 q_1 + r^4 p_1^6 q_1^4 + \ldots)$$
$$= a_0 \sum_{s=0}^{\infty} r^s p_1^{\frac{s(s-1)}{1.2}} q_1^{\frac{s(s-1)(s-2)}{1.2.3}}.$$

Write $q_1 = 1 + q$. Generally q is very much smaller than p so that q^2 and higher powers may be safely neglected.

$$S_3 = a_0 [1 + r + r^2 p_1 + r^3 p_1^3 (1 + q) + r^4 p_1^6 (1 + q)^4 + \ldots]$$
$$\cong a_0 [(1 + r + r^2 p_1 + r^3 p_1^3 + r^4 p_1^6 + \ldots)$$
$$+ q r^3 p_1^3 (1 + 4 r p_1^3 + 10 r^2 p_1^7 + 20 r^3 p_1^{12} + \ldots)].$$

The sign $\cong$ denotes "approximately equal to." We will now obtain an approximation to the sum of the series

$$S' = 1 + 4 r p_1^3 + 10 r^2 p_1^7 + 20 r^3 p_1^{12} + \ldots.$$

p_1 is less than unity because the convergence of the tail of the hypergeometric series is more rapid than that of the geometrical $a_0 (1 + r + r^2 + r^3 + \ldots)$ whose first two terms agree with those of the tail of the hyperseries. Hence

$$S' < 1 + 4r + 10r + \ldots = (1 - r)^{-4}.$$

Moreover,

$$S' < 1 + 4 r p_1^3 + 10 r^2 p_1^6 + 20 r^3 p_1^9 + \ldots = (1 - r p_1^3)^{-4} < (1 - r)^{-4}.$$

Hence $(1 - r p_1^3)^{-4}$ is a better upper limit to the sum S' than $(1 - r)^{-4}$. There is, however, very little difference between these two expressions because p_1 is close to unity.

When $p = (1 - p_1)$ is sufficiently small, say of the order ·01, we may obtain a still closer approximation and, incidentally, a lower limit to S'.

$$S' = 1 + 4 r p_1^3 + 10 r^2 p_1^7 + 20 r^3 p_1^{12} + \ldots$$
$$\cong 1 + 4r (1 - 3p) + 10 r^2 (1 - 7p) + 20 r^3 (1 - 12p) + \ldots$$
$$= (1 + 4r + 10 r^2 + 20 r^3 + \ldots) - rp (12 + 70 r + 240 r^2 + \ldots)$$
$$= [(1 - r)^{-4} - 2 rp (6 - r) (1 - r)^{-6}].$$

Evidently

$$(1 - r p_1^3)^{-4} > S' > [(1 - r)^{-4} - 2 rp (6 - r) (1 - r)^{-6}].$$

Hence *a third approximation* to the sum of the tail is

$$S_3 = a_0 (1 - r)^{-1} [[1 - p r^2 (1 - r)^{-2} + 3 p^2 r^3 (1 - r)^{-4}$$
$$- r^3 p^3 (1 + 13 r + r^2) (1 - r)^{-6}\}$$
$$+ q r^3 p_1^3 (1 - r)^{-3} \{1 - 2 pr (6 - r) (1 - r)^{-2}\}],$$

where $$p = 1 - p_1, \quad q = q_1 - 1.$$

Examples.

I. *Sum the tail of the series* $F(30, 30, 111, \cdot75)$ *starting with the* 28*th term.*

The tail of the series is

$$\cdot796,454 + \cdot502,267 + \cdot314,369 + \ldots.$$

We readily find

$$\begin{aligned} r_1 &= \cdot6306,2901, \\ p_1 &= \cdot9925,014, \qquad & p &= \cdot007,4986, \\ q_1 &= 1\cdot000,64443, & q &= \cdot000,64443, \\ u_1 &= \cdot999,93951, & 1 - u_1 = u &= \cdot000,06049, \\ &\ldots\ldots\ldots\ldots & \ldots\ldots & \qquad \ldots\ldots\ldots\ldots \end{aligned}$$

The true sum of the tail found by computing each term separately is 2·117,841. Further,

$$\begin{aligned} S_1 &= 2\cdot156,244, \\ S_2 &= 2\cdot156,244\,(1 - \cdot021,8576 + \cdot002,27275 - \cdot000,39944) \\ &= 2\cdot113,153, \\ S_3 &= 2\cdot113,153 + \cdot004,2444 \\ &= 2\cdot117,397. \end{aligned}$$

This differs from the true value by only 4 in the fifth place.

II. *Sum the tail of the series* $F(50, 50, 100, \frac{1}{2})$ *starting with the* 81*st term.*

The tail of the series is

$$a_0\,(1 + \cdot578,123 + \cdot333,418 + \cdot191,836 + \ldots).$$

Its true sum is $2\cdot360,158a_0$.

We readily find

$$\begin{aligned} r &= \cdot578,12289, \\ p_1 &= \cdot997,581112, \qquad & p &= \cdot002,418888, \\ q_1 &= 1\cdot0000,604352, & q &= \cdot0000,604352, \\ S_1 &= 2\cdot370,3585a_0, \\ S_2 &= 2\cdot370,3585a_0\,(1 - \cdot004,54239 + \cdot0000,7138 - \cdot00000,429) \\ &= 2\cdot359,7505a_0, \\ S_3 &= (2\cdot359,7505 + \cdot000,33479)\,a_0 \\ &= 2\cdot360,085a_0. \end{aligned}$$

The last approximation differs from the true value by 7 in the sixth place.

III. *Sum the tail of the series* $F(50, 50, 100, \frac{1}{2})$ *starting with the* 61*st term.*

The tail of the series is

$$S = a_0'\,(1 + \cdot617,161 + \cdot379,270 + \cdot232,118 + \ldots).$$

True value of S is $2\cdot585,789a_0'$.

We find the following values for the constants:

$$r = \cdot 617,160890,$$
$$p_1 = \cdot 995,7540,680, \qquad p = \cdot 004,245932,$$
$$q_1 = 1 \cdot 000,13065, \qquad q = \cdot 000,13065,$$

and
$$S_1 = 2 \cdot 612,0623 a_0',$$
$$S_2 = 2 \cdot 612,0623 a_0 (1 - \cdot 011,03411 + \cdot 000,59183 - \cdot 0000,2058)$$
$$= 2 \cdot 584,7335 a_0',$$
$$S_3 = (2 \cdot 584,7335 + \cdot 001,1399) a_0'$$
$$= 2 \cdot 585,873 a_0'.$$

S_3 differs from the true value by 1 in the fifth place.

IV. *Sum the tail of the series* $F(-100, -100, 1, \frac{1}{2})$ *starting with the* 10*th term after the mode.*

Modal term is the 42nd. We find

$$r = \cdot 443,97189,$$
$$p_1 = \cdot 923,73046, \qquad p = \cdot 076,26954,$$
$$q_1 = 1 \cdot 000,156141, \qquad q = \cdot 000,156141.$$

True sum of tail is $\cdot 751,702$ multiplied by a constant which will be dropped in the subsequent work.

$$S_1 = \cdot 781,4122,$$
$$S_2 = \cdot 781,4122 (1 - \cdot 048,626 + \cdot 015,9773 - \cdot 005,09085).$$

p is too large to apply formula S_2 with any great degree of accuracy. It is interesting, however, to examine how closely the successive approximations to S_2 represent the true value of S. These approximations are

(i) $\cdot 781,412$, (ii) $\cdot 743,415$,
(iii) $\cdot 755,900$, (iv) $\cdot 748,746$.

The formulae S_2 is not entirely satisfactory when p is larger than $\cdot 01$. For such cases, we may obtain a very close approximation to S_2 by the application of the Euler-Maclaurin theorem to the summing of the series

$$\sum_{s=0}^{\infty} r^s p_1^{\frac{s(s-1)}{1.2}}.$$

The problem, however, will be approached from a wider angle, namely, by applying the Euler-Maclaurin theorem to the summing of the more general series

$$\sum_{s=0}^{\infty} r^s p_1^{\frac{s(s-1)}{1.2}} q_1^{\frac{s(s-1)(s-2)}{1.2.3}} u_1^{\frac{s(s-1)(s-2)(s-3)}{1.2.3.4}} \ldots.$$

§ 8. *Application of the Euler-Maclaurin Theorem to the summing of the Series*

$$S = a_0 \sum_{s=0}^{\infty} r_1^s p_1^{\frac{s(s-1)}{1.2}} q_1^{\frac{s(s-1)(s-2)}{1.2.3}} u_1^{\frac{s(s-1)(s-2)(s-3)}{1.2.3.4}} \ldots.$$

Write $\log_e r_1 = a,$

$\log_e p_1 = b,$

$\log_e q_1 = c,$ etc.,

then $$S = a_0 \sum_{s=0}^{\infty} e^{as + b\frac{s(s-1)}{1.2} + c\frac{s(s-1)(s-2)}{1.2.3} + \dots} \quad \dots\dots\dots\dots\dots\text{(i)}.$$

The quantities $a, b, c, \dots$ generally tend rapidly to zero. Clearly,

$$S = a_0 \sum_{s=0}^{\infty} e^{as + b\frac{s(s-1)}{2}} \left[1 + \left\{c\frac{s(s-1)(s-2)}{1.2.3} + d\frac{s(s-1)(s-2)(s-3)}{1.2.3.4} + \dots\right\} + \frac{1}{2}\left\{c\frac{s(s-1)(s-2)}{1.2.3} + d\frac{s(s-1)(s-2)(s-3)}{1.2.3.4} + \dots\right\}^2 + \dots\right] \quad \dots\text{(ii)}$$

$$= a_0 \sum_{s=0}^{\infty} e^{as + b\frac{s(s-1)}{2}} [1 + c_0 s + c_1 s^2 + c_2 s^3 + \dots] \quad \dots\dots\dots\dots\dots\text{(iii)}$$

$$= a_0 \Sigma e^{-c_2' s^2 - c_1' s} [1 + c_0 s + c_1 s^2 + c_2 s^3 + \dots] \quad \dots\dots\dots\dots\dots\text{(iv)},$$

where $c_2' = -b/2, \quad c_1' = -(a - \frac{1}{2}b)$

$$= a_0 \sum_{s=0}^{\infty} e^{\frac{1}{2}x_1^2} e^{-\frac{1}{2\sigma^2}(s + x_1\sigma)^2} [1 + c_0 s + c_1 s^2 + \dots],$$

where $$x_1 = c_1'\sigma, \quad \sigma = \frac{1}{\sqrt{(2c_2')}}.$$

Hence, by the Euler-Maclaurin theorem,

$$S \cong a_0 e^{\frac{1}{2}x_1^2}\left[\int_0^{\infty} e^{-\frac{1}{2\sigma^2}(x + x_1\sigma)^2} \{1 + c_0 x + c_1 x^2 + \dots\}\, dx + E_s\right]$$

$$= a_0 e^{\frac{1}{2}x_1^2}\left[\sigma \int_{x_1}^{\infty} e^{-\frac{1}{2}x^2} \{1 + \xi_1 x + \xi_2 x^2 + \dots\}\, dx + E_s\right] \quad \dots\dots\dots\dots\text{(iv}b\text{)},$$

where E_s is the Euler-Maclaurin correction *.

* We may also write

$$S \cong a_0 e^{\frac{1}{2}x_1^2}\left[\sigma \int_{\left(x_1 - \frac{1}{2\sigma}\right)}^{\infty} e^{-\frac{x^2}{2}} \{1 + \xi_1 x + \xi_2 x^2 + \dots\}\, dx + E_s'\right],$$

where E_s' is approximately equal to

$$\cdot 0387{,}153 a_3 - \cdot 1578{,}127 a_2 + \cdot 2411{,}459 a_1 - \cdot 1220{,}458 a_0.$$ (See § 10, p. 106.)

More accurate expressions for E_s' are the following:

${}_6E_s'$	${}_8E_s'$
$\frac{1}{2}a_0 + 0{\cdot}195{,}7755 a_0$	$\frac{1}{2}a_0 + 0{\cdot}213{,}0246 a_0$
$- 0{\cdot}460{,}3838 a_1$	$- 0{\cdot}589{,}0198 a_1$
$+ 0{\cdot}546{,}5364 a_2$	$+ 0{\cdot}964{,}0148 a_2$
$- 0{\cdot}471{,}4289 a_3$	$- 1{\cdot}240{,}8904 a_3$
$+ 0{\cdot}260{,}6070 a_4$	$+ 1{\cdot}140{,}5644 a_4$
$- 0{\cdot}082{,}4736 a_5$	$- 0{\cdot}720{,}9438 a_5$
$+ 0{\cdot}011{,}3674 a_6$	$+ 0{\cdot}297{,}8547 a_6$
	$- 0{\cdot}072{,}4970 a_7$
	$+ 0{\cdot}007{,}8926 a_8$

These last two expressions are obtained from a form of the Euler-Maclaurin theorem due to Professor Karl Pearson, the former based on sixth differences and the latter on eighth differences.

The integral $\int_{x_1}^{\infty} e^{-\frac{1}{2}x^2} x^s dx$ is the sth incomplete normal moment function and has been tabulated up to the twelfth. S can, therefore, be expressed in terms of incomplete normal moment functions.

As is frequently the case, the constants c, d, e, ... are sufficiently small to be neglected altogether. Then

$$S \cong a_0 e^{\frac{1}{2}x_1^2} \left(\sigma \int_{x_1}^{\infty} e^{-\frac{1}{2}x^2} dx + E_1 \right) \quad \text{.....(v).}$$

This is, in effect, the expression obtained by Professor Burton H. Camp* for the sum of the tail of a discrete frequency distribution. He approached the problem differently, namely, by fitting the tail of a normal curve to the tail of the distribution. (v) has been put in the convenient form†

$$S \cong a_0 \left\{ \frac{1 - a_{x_1}}{z_{x_1}} \sigma + \psi \right\} \quad \text{.....(vi),}$$

where

$$\psi = \frac{1}{2} + \frac{1}{12} c_1 \left[1 + \frac{3 - x_1^2}{60\sigma^2} + \frac{15 - 10x_1^2 + x_1^4}{2520\sigma^4} \right] \quad \text{.....(vi } b\text{).}$$

This formula has been applied with a large measure of success to the summing of the tail of a binomial and hypergeometric series in which the last element x is unity. It fails, however, when the stump lies within 2σ of the mode, and is only reliable for points outside mode $\pm 3\sigma$. Similarly, when applied to the general hypergeometric series, the results are reliable only when the stump lies outside the range mode $\pm 3\sigma$. When the stump lies within mode $\pm 3\sigma$ it may be found necessary to include a few terms in the expansion (iv b). The formula (vi) is essentially equivalent to the approximation S_2 obtained for the series

$$a_0 \sum_{s}^{0,\infty} r_1^s p_1^{\frac{s(s-1)}{2}}.$$

When, as is very frequently the case, $p = 1 - p_1$ is not appreciably greater than ·01, S_2 holds several advantages over formula (vi); it may be applied more easily, the contribution of each successive term is made clear, and, since S_2 is concerned only with the fitting of a series to another, no correction such as the Euler-Maclaurin need be applied.

Retaining the constant c in (i), we find

$$S \cong a_0 \sum_{s=0}^{\infty} e^{\frac{1}{2}x_1^2} e^{-\frac{1}{2\sigma^2}(s + x_1\sigma)^2} \left\{ 1 + c \frac{s(s-1)(s-2)}{1.2.3} \right\}.$$

Note that $$c = \log_e q_1 = \log_e (1+q) = q - \frac{q^2}{2} + \frac{q^3}{3} - \ldots.$$

Therefore $$\left\{ 1 + c \frac{s(s-1)(s-2)}{1.2.3} \right\} < \left\{ 1 + q \frac{s(s-1)(s-2)}{1.2.3} \right\}.$$

Moreover, $$(1+q)^{\frac{s(s-1)(s-2)}{1.2.3}} > 1 + \frac{s(s-1)(s-2)}{1.2.3} \cdot q.$$

* *Biometrika*, Vol. XVI. p. 163.

† *Tables for Statisticians and Biometricians*, Part II, Introduction, pp. xxx—xl.

Hence, a better approximation to S is

$$a_0 e^{\frac{1}{2}x_1^2} \sum_{s=0}^{\infty} e^{-\frac{1}{2\sigma^2}(s+x_1\sigma)^2} \left\{1 + q\frac{s(s-1)(s-2)}{1.2.3}\right\}.$$

This differs from (vi) by the quantity

$$M = a_0 e^{\frac{1}{2}x_1^2} \sum_{s=0}^{\infty} \left[e^{-\frac{1}{2\sigma^2}(s+x_1\sigma)^2} q \cdot \frac{s(s-1)(s-2)}{1.2.3}\right] = a_0 e^{\frac{1}{2}x_1^2} \sum_{s=3}^{\infty} [\text{same expression}].$$

This is essentially the quantity added to S_2 (p. 94) to obtain the closer approximation S_3. It is identical with the series

$$a_0 q r^3 p_1^3 (1 + 4rp_1^3 + 10r^2 p_1^7 + 20r^3 p_1^{12} + \ldots).$$

M accordingly satisfies the inequalities

$$a_0 q r^3 p_1^3 (1-r)^{-4} > a_0 q r^3 p_1^3 (1 - rp_1^3)^{-4} > M$$

$$> a_0 q r^3 p_1^3 (1-r)^{-4} \left\{1 - \frac{2r(1-p_1)(6-r)}{(1-r)^2}\right\} \quad \ldots\ldots\ldots\ldots\ldots\text{(vii)}.$$

These inequalities give a good indication of the accuracy of the result obtained by the application of Camp's formula (vi) and whether it is necessary to go a stage further in the approximation.

M may also be expressed in the form

$$a_0 q \left[e^{\frac{1}{2}x_1^2} \sigma \int_{\frac{3}{\sigma}+x_1}^{\infty} e^{-\frac{1}{2}x^2} \{\sigma(x-x_1)\} \{\sigma(x-x_1)-1\} \{\sigma(x-x_1)-2\}\, dx + E\right] \text{(viii)}.$$

If $$t_s = e^{as+bs(s-1)/2} \cdot \frac{s(s-1)(s-2)}{6},$$

then $$t_3 = r^3 p_1^3,$$

$$t_3' = r^3 p_1^3 \left(\frac{11}{6} + a + \frac{5b}{2}\right),$$

$$t_3''' = r^3 p_1^3 \left(1 + 6a + \frac{41b}{2}\right),$$

neglecting squares and products of a and b.

Hence $$E \cong r^3 p_1^3 \left[\frac{1}{2} + \frac{11 + 6a + 15b}{72} - \frac{2 + 12a + 41b}{1440}\right].$$

Since q is itself small, one term only of E generally suffices. The evaluation of (viii) is often a little troublesome. However, in most cases the right-hand side of inequality (vii) gives sufficiently accurate results.

The most satisfactory way so far, then, of finding the sum of a tail of a hypergeometric series when the "stump" does not lie within the mode $\pm 2{\cdot}5\sigma$ is to

(a) obtain the approximate sum by the application of formula (vi),

(b) correct this result, when necessary, by evaluating M.

When $p = (1 - p_1)$ is not appreciably larger than ·01, the right-hand side of inequality (vii) may be taken as a good approximation to M. In all cases, however, M is less than $a_0 q r^3 p_1^3 (1 - rp_1^3)^{-4}$.

Examples.

I. (See above.) *Sum the tail of* $F(30, 30, 111, \cdot 75)$ *starting with the* 28*th term.*

The approximation obtained by using (vi) is 2·113,316.

The correction (taking into account the value of q) obtained from the right-hand side of inequality (vii) is ·004,2444. Hence,

$$\left.\begin{array}{l}\text{Corrected sum} = 2\cdot 117{,}560\\ \text{True sum} \quad\;\; = 2\cdot 117{,}841\end{array}\right\}.$$

II. *Sum the tail of* $F(50, 50, 100, \frac{1}{2})$ *starting with the* 81*st term.*

$$\left.\begin{array}{ll}\text{Formula (vi) gives} & 2\cdot 359{,}836,\\ \text{Correction} & = \cdot 000{,}335,\\ \text{Corrected sum} & = 2\cdot 360{,}171\\ \text{True sum} & = 2\cdot 360{,}158\end{array}\right\}.$$

III. *Sum the tail of* $F(50, 50, 100, \frac{1}{2})$ *starting with the* 61*st term.*

$$\left.\begin{array}{ll}\text{Formula (vi) gives} & 2\cdot 584{,}6630,\\ \text{Correction from (vii)} & = \cdot 001{,}1399,\\ \text{Corrected sum} & = 2\cdot 585{,}803\\ \text{True sum} & = 2\cdot 585{,}789\end{array}\right\}.$$

IV. *Sum the tail of* $F(-100, -100, 1, \frac{1}{2})$ *starting with the* 52*nd term.*

$$\left.\begin{array}{ll}\text{Formula (vi) gives} & \cdot 751{,}738\\ \text{True sum is} & \cdot 751{,}702\end{array}\right\}.$$

In this case it is not necessary to apply any correction, formula (vi) gives sufficient accuracy. The result cannot be trusted in the fifth place because ψ does not converge sufficiently rapidly for the approximate value (vi *b*) to give five or more figure accuracy. Applying the left-hand side of inequality (vii) we readily find that the correction is less than ·0000,3325, i.e. the result differs from the true value by less than 3·3 in the fifth place. The right-hand side of (vii) does not give a closer result because $p = \cdot 07$ is rather large.

§ 9. *On the Fitting of the Tail of a Type III Curve to the Tail of a Frequency Distribution.*

The above examples shew clearly that Camp's method for the evaluation of the tail of a discrete frequency distribution by fitting to it the tail of a normal curve is not always adequate, especially in those cases in which the third ratios q_i differ appreciably from unity. In general, the correction given by formula (vii) makes a considerable difference in the results; the accuracy being improved by one, two and sometimes three figures.

The fitting of the tail of a normal curve to the tail of a frequency distribution $F(z)$ depends on the evaluation of three constants. We can, alternatively, fit the tail of the Type III curve

$$y = y_0 e^{-\rho z}(a + z)^t. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(i)}.$$

This will depend on the evaluation of four independent constants and, accordingly, we should expect a certain amount of improvement in the results.

The cases, $F(z)$ continuous and $F(z)$ discontinuous, will be treated separately.

$F(z)$ continuous.

The problem will be approached by fitting (i) to the tail $z \geqslant z_0$ of $F(z)$ by equating the corresponding logarithmic differentials at the stump.

Let $u = \log F(z)$ and $a_s = \dfrac{d^s}{dz^s} F(z)$ at $z = z_0$. Equate these to the logarithmic differentials of (i) at $z = 0$. There result the following equations, which have to be solved for the constants of the Type III curve:

$$\begin{aligned} -\rho + t/a &= a_1, \\ -t/a^2 &= a_2, \\ 2t/a^3 &= a_3. \end{aligned}$$

Clearly,

$$\begin{aligned} a &= -2a_2/a_3, \\ t &= -4a_2^3/a_3^2, \\ \rho &= (2a_2^2/a_3 - a_1). \end{aligned}$$

y_0 may be found by equating the ordinates at the respective stumps.

Area of the tail $z \geqslant z_0$ of $F(z)$

$$\begin{aligned} &= \frac{1}{a^t} F(z_0) \int_0^\infty e^{-\rho z} (a+z)^t \, dz \\ &= \frac{1}{a^t} F(z_0)\, e^{\rho a} \int_a^\infty e^{-\rho z} z^t \, dz \\ &= F(z_0) \frac{\alpha_a}{y_a}, \end{aligned}$$

where α_a is the area of the tail $z \geqslant a$ of the Type III curve $y = y_0 e^{-\rho z} z^t$ and y_a the ordinate at $z = a$. The quantity α_a may be found from the *Tables of Incomplete* Γ-*Function.*

$F(z)$ discontinuous $\equiv (a_0, a_1, a_2, \ldots)$.

I. Consider the discrete terms a_i $(i = 0, 1, 2, \ldots)$ as being the ordinates of a continuous function $f(z)$ at unit intervals of the argument z, the first term a_0 being given by $z = 0$.

Let a_s be the first term of the tail of $F(z)$. We require the first three logarithmic differentials of $f(z)$ at $z = s$. These may be found as follows. Calculate (to ten figures) the natural logarithms of the first six terms of the tail, calculate the differences and substitute in either of the following sets of formulae giving the differentials in terms of the differences.

*Forward Difference Formulae**:

$$\begin{aligned} hf_0' &= (\Delta - \tfrac{1}{2}\Delta^2 + \tfrac{1}{3}\Delta^3 - \tfrac{1}{4}\Delta^4 + \tfrac{1}{5}\Delta^5 - \tfrac{1}{6}\Delta^6 + \ldots) f_0, \\ h^2 f_0'' &= (\Delta^2 - \Delta^3 + \tfrac{11}{12}\Delta^4 - \tfrac{5}{6}\Delta^5 + \tfrac{137}{180}\Delta^6 - \ldots) f_0, \\ h^3 f_0''' &= (\Delta^3 - \tfrac{3}{2}\Delta^4 + \tfrac{7}{4}\Delta^5 - \tfrac{15}{8}\Delta^6 + \ldots) f_0. \end{aligned}$$

* Whitaker and Robinson, *Calculus of Observations.*

Central Difference Formulae:*

$$hf_0' = \tfrac{1}{2}(f_1 - f_{-1}) - \tfrac{1}{12}(\delta^2 f_1 - \delta^2 f_{-1}) + \tfrac{1}{60}(\delta^4 f_1 - \delta^4 f_{-1}) - \tfrac{1}{280}(\delta^6 f_1 - \delta^6 f_{-1}) + \dots,$$

$$h^2 f_0'' = \delta^2 f_0 - \tfrac{1}{12}\delta^4 f_0 + \tfrac{1}{90}\delta^6 f_0 - \tfrac{1}{560}\delta^8 f_0 + \dots,$$

$$h^3 f_0''' = \tfrac{1}{2}(\delta^2 f_1 - \delta^2 f_{-1}) - \tfrac{1}{8}(\delta^4 f_1 - \delta^4 f_{-1}) + \tfrac{7}{240}(\delta^6 f_1 - \delta^6 f_{-1}) + \dots.$$

The constants of the Type III curve are then found as in the previous case.

If S is the sum of the tail of $F(z)$, then

$$S \cong \left[y_0 \int_0^\infty e^{-\rho z}(a+z)^t + E \right],$$

where E is the Euler-Maclaurin correction†. Furthermore,

$$S \cong a_a \left[a^{-t} e^{a\rho} \int_a^\infty e^{-\rho z} z^t \, dz + \psi \right] \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\text{(ii)},$$

where

$$\psi = \frac{1}{2} + \frac{1}{12a}\left[(t-\rho a) - \frac{1}{60a^2}\{t(t-1)(t-2) - 3t(t-1)a\rho + 3ta^2\rho^2 - a^3\rho^3\} \right.$$
$$+ \frac{1}{2520a^4}\{t(t-1)(t-2)(t-3)(t-4) - 5t(t-1)(t-2)(t-3)a\rho$$
$$\left. + 10t(t-1)(t-2)a^2\rho^2 - 10t(t-1)a^3\rho^3 + 5ta^4\rho^4 - a^5\rho^5\} - \dots \right].$$

If a_a is the area of the tail of the Type III curve $y = y_0 e^{-\rho z} z^t$, $z \geqslant a$, and z_a the ordinate at $z = a$, S may be thrown into the convenient form

$$S \cong a_a \left(\frac{a_a}{z_a} + \psi \right).$$

II. Alternatively, we may find the Type III curve which passes through the first four terms of the tail of $F(z)$. Thus, let $y = y_0 e^{-\rho z}(a+z)^t$ be the equation of a Type III curve, and let d_1, d_2, d_3 be the first three logarithmic forward differences of the first term, a_a, of the tail of $F(z)$. The logarithmic differences of y at $z = 0$ are, respectively,

1st	$-\rho + t\log(a+1)/a,$
2nd	$t\log(a+2)\,a/(a+1)^2,$
3rd	$t\log(a+3)\,(a+1)^3/(a+2)^3 a.$

Equate these to the corresponding differences of a_a. This results in three equations to solve for the constants ρ, a and t. Eliminate t from the last two equations and we have

$$d_2 \log \frac{(a+3)(a+1)^3}{(a+2)^3 a} = d_3 \log \frac{(a+2)\,a}{(a+1)^2},$$

i.e. $$d_2 \log(a+3) - (3d_2 + d_3)\log(a+2)$$
$$+ (3d_2 + 2d_3)\log(a+1) - (d_2 + d_3)\log a = 0 \{\equiv X(a)\}.$$

* Sheppard, *Proc. Lond. Math. Soc.* Vol. xxxi. p. 465.

† The correction E may, perhaps, be obtained more readily by using the formulae in the footnote, p. 97.

This equation may be solved quite readily by forming a small table of the values of the function $X(a)$ and interpolating for the value of a which makes X zero.

Having found a, the other constants may be determined by simple substitution. The sum S is then given by formula (ii) above.

Examples *.

I. *Fit a Type III curve to the tail of* $F(50, 50, 100, \frac{1}{2})$ *starting with the* 81*st term.*

The first six forward differences of the common logarithms of the first term of the tail are, respectively,

$$\Delta = -\cdot 237,9798,316, \qquad \Delta^2 = -\cdot 00105,17821,$$
$$\Delta^3 = +\cdot 000,0263,063, \qquad \Delta^4 = -\cdot 00000,08886,$$
$$\Delta^5 = +\cdot 000,0000,383, \qquad \Delta^6 = -\cdot 00000,00021.$$

Converting into natural logarithms and substituting in the forward difference formulae, the first three logarithmic differentials reduce to

$$D = -\cdot 546,737,414,$$
$$D^2 = -\cdot 002,484,344,$$
$$D^3 = \cdot 000,0638,021.$$

The Type III curve which has the same differentials at $z = 0$, is

$$y = y_0 e^{-\cdot 740,2096z} (77\cdot 876,5589 + z)^{15\cdot 066,946} \quad \text{.....(iii).}$$

y_0 is found by equating y (at $z = 0$) to the first term of the tail of the series. Hence $y_0 = (77\cdot 876,559)^{-15\cdot 066,946}$.

Terms of tail	Ordinates of (i)	Terms of tail	Ordinates of (i)
1·000,000	1·000,000	·000,678	·000,678
·578,123	·578,123	·000,381	·000,381
·333,418	·333,417	·000,214	·000,214
·191,836	·191,836	·000,120	·000,120
·110,121	·110,122	·000,067	·000,067
·063,073	·063,072	·000,038	·000,038
·036,046	·036,046	·000,021	·000,021
·020,556	·020,556	·000,012	·000,012
·011,698	·011,698	·000,006	·000,006
·006,644	·006,644	·000,004	·000,004
·003,766	·003,766	·000,002	·000,002
·002,130	·002,130	·000,001	·000,001
·001,203	·001,203		

The fit of the Type III curve to the tail of the series is surprisingly good, and the area of the tail, corrected by the Euler-Maclaurin theorem, agrees to six figures with the true sum of the tail. This is a decided improvement on any of the previous methods.

* In the first two examples the actual terms of the series have been multiplied throughout by a constant to reduce the first term of the tail to unity.

II. *Fit a Type III curve to the tail of* $F(50, 50, 100, \frac{1}{2})$ *starting with the* 61*st term.*

Using the forward difference formula, the first three logarithmic differentials are found to be

$$D = -\cdot 480{,}453{,}0646, \qquad D^2 = -\cdot 004{,}390{,}8871,$$
$$D^3 = \cdot 000{,}139{,}3941,$$

and the required Type III curve is

$$y = y_0 e^{-\cdot 757{,}0772z}(z + 62\cdot 999{,}6028)^{17\cdot 427{,}2117} \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(iv)},$$

where $$y_0 = (62\cdot 999{,}6028)^{-17\cdot 427{,}2117}.$$

Terms of tail	Ordinates of (i)	Terms of tail	Ordinates of (i)
1·000,000	1·000,000	·000,824	·000,824
·617,161	·617,161	·000,484	·000,484
·379,270	·379,270	·000,284	·000,283
·232,118	·232,117	·000,166	·000,165
·141,492	·141,491	·000,096	·000,096
·085,915	·085,914	·000,056	·000,056
·051,972	·051,971	·000,032	·000,031
·031,325	·031,324	·000,019	·000,018
·018,813	·018,812	·000,011	·000,010
·011,260	·011,259	·000,006	·000,006
·006,117	·006,116	·000,004	·000,004
·003,994	·003,993	·000,002	·000,002
·002,367	·002,366	·000,001	·000,001
·001,399	·001,398		

Here again the fit is surprisingly close. The corrected area of the tail of the Type III curve agrees to five figures with the true sum of the series. It is out by only 2 in the sixth place.

III. *Fit a Type* III *curve to the tail of* $F(30, 30, 111, \frac{3}{4})$ *starting with the* 28*th term.*

The first three forward differences of the common logarithm of the first term are, respectively,

$$d_1 = -\cdot 200{,}2261{,}429, \qquad d_2 = -\cdot 003{,}2694{,}047,$$
$$d_3 = \cdot 000{,}2809{,}485.$$

The equation of the Type III curve which passes through the first four terms of the tail is readily found to be

$$y = y_0 e^{-\cdot 628{,}746z}(20\cdot 78533 + z)^{3\cdot 569{,}066} \quad \ldots\ldots\ldots\ldots\ldots\ldots\text{(v)},$$

where $y_0 = \cdot 796{,}454/y$ (at $z = 0$).

The true sum of the tail of the series is 2·117,841 and the area of the tail of the Type III curve (i) corrected by the Euler-Maclaurin theorem is 2·117,662. The fit of the curve to the series is not so good as in the previous two examples, but still, it is an improvement on any of the previous methods.

Terms of tail	Ordinates of (i)	Terms of tail	Ordinates of (i)
·796,454	·796,454	·001,283	·001,272
·502,267	·502,267	·000,762	·000,752
·314,369	·314,369	·000,451	·000,444
·195,414	·195,414	·000,267	·000,261
·120,708	·120,706	·000,157	·000,153
·074,135	·074,127	·000,093	·000,090
·045,291	·045,279	·000,054	·000,052
·027,537	·027,520	·000,032	·000,030
·016,669	·016,649	·000,019	·000,018
·010,049	·010,029	·000,011	·000,011
·006,036	·006,017	·000,006	·000,006
·003,614	·003,597	·000,002	·000,002
·002,157	·002,142	·000,001	·000,001

§ 10. It was shown in § 1 that when the number of significant terms of the hypergeometric series $F(\alpha, \beta, \gamma, x)$ is not small, the curve for which $\frac{1}{y}\frac{dy}{dz}$ is equal to $\Delta y_r/\{\frac{1}{2}c\,(y_r + y_{r+1})\}$, $r = 1, 2, 3, \ldots$, fits the series quite well around the mode but may deviate a little towards the extreme tails. The probability integral of the series may be replaced by that of a continuous curve of the type $G(z) = y_0 e^{-\rho z} P(z)$ with sufficient accuracy for most statistical purposes.

An examination of the tables in the latter part of § 1 will reveal that the ordinates of the curve $G(z)$ correspond more closely to the terms of the series than the areas under the curve. If, therefore, the constant of integration, y_0, is calculated by equating the maximum term of the series to the corresponding mid-ordinate of the curve, the mid-ordinates around the mode will agree even more closely with the terms of the series. The deviation towards the tails, however, will be more pronounced. If, therefore, a more accurate determination of the terms *around the mode* is required, the following method may be used:

Let $G(z)$ be the curve obtained from the differential equation $\frac{1}{y}\frac{dy}{dz} = \frac{2}{c}\frac{\Delta y_r}{(y_r + y_{r+1})}$, the constant of integration being determined by equating the modal term of the series to the corresponding mid-ordinate of the curve (note that this is not necessarily the mode of $G(z)$). The approximate value of the sum of a number of terms near the mode is, therefore,

$$S = \int_{z_1}^{z_2} G(z)\,dz + E,$$

where z_1 and z_2 are the appropriate limits and E the Euler-Maclaurin correction. A close approximation to E may be found as follows:

If y_r is the $(r+1)$st mid-ordinate of the curve, the area under this part of the curve is given approximately by

$$\frac{1}{5760}[5178 y_r + 308\,(y_{r+1} + y_{r-1}) - 17\,(y_{r+2} + y_{r-2})];$$

the grouping unit c is taken to be unity.

Hence, if we wish to sum the terms from the sth to the tth,

$$\int_s^t G(z)\,dz = \sum_{r=s+1}^{t} y_r - \cdot0387{,}153\,(y_{s+4}+y_{t-3}) + \cdot1578{,}127\,(y_{s+3}+y_{t-2}) - \cdot2411{,}459\,(y_{s+2}+y_{t-1}) + \cdot1220{,}485\,(y_{s+1}+y_t).$$

Therefore,

$$\sum_{r=s+1}^{t} y_r = \int_s^t G(z)\,dz + \cdot0387{,}153\,(y_{s+4}+y_{t-3}) - \cdot1578{,}127\,(y_{s+3}+y_{t-2}) + \cdot2411{,}459\,(y_{s+2}+y_{t-1}) - \cdot1220{,}485\,(y_{s+1}+y_t).$$

The terms to the right of the integral make up the correction. Σy_r is approximately equal to the required sum of the terms of the series near the mode.

When the fourth element x is not small, the method given in §2 for finding the partial sums of $F(\alpha, \beta, \gamma, x)$ is quite satisfactory for all parts of the series except the start. The curve deduced by this method fits the series very closely at points beyond the mode but may deviate appreciably at the start. The range of disagreement extends rapidly towards the mode as x becomes small.

The particular method to adopt for finding the sum of a number of terms, when large, of the hypergeometric series depends on the type of series and the position of the terms relative to the mode. It will be sufficient to consider sums of the type S_t (= sum of the first t terms) because the sum of any number of successive terms may be expressed as a difference of two S's.

1. When the last term of S_t (i.e. the tth) falls outside the range mode $\pm 2{\cdot}5\sigma$, one of the methods of §§7—9 may be used for evaluating S_t. The list of methods in order of accuracy is as follows:

(i) Fit a geometrical series $a_0 \sum_{s=0}^{\infty} r^s$ to the tail and obtain the approximation S_1 (p. 93).

(ii *a*) Fit the series $a_0 \sum_{s=0}^{\infty} r^s p_1^{s(s-1)/2}$ and obtain the approximation S_2 (p. 93).

(ii *b*) Use the equivalent method of p. 98 (Camp's method).

(iii *a*) Fit the series $a_0 \sum_{s=0}^{\infty} r^s p_1^{\frac{s(s-1)}{1.2}} q_1^{\frac{s(s-1)(s-2)}{1.2.3}}$ and obtain the approximation S_3 (p. 94).

(iii *b*) Correct the result obtained by (iv *b*) as in p. 97.

(iv) Fit a Type III curve to the tail of the series by either

(*a*) equating the logarithmic differences at the stumps, or

(*b*) equating the logarithmic differentials at the stumps.

The required sum may then be expressed by an incomplete Gamma function (pp. 103—104).

(v) Any degree of accuracy may be obtained by taking a sufficient number of constants in the series $a_0 \sum_{s=0}^{\infty} r^s \, p_1^{\frac{s(s-1)}{1.2}} \, q_1^{\frac{s(s-1)(s-2)}{1.2.3}} \, u_1^{\frac{s(s-1)(s-2)(s-3)}{1.2.3.4}} \ldots$ and expressing the sum of the tail in terms of incomplete normal moment functions.

2. When the tth term lies near the mode, then for

(i) x *small* ($< \frac{1}{3}$), use method of § 1,

(ii) $x \geqslant \frac{1}{3}$, use method of § 2.

3. For J-shaped series, use method developed in § 2 whatever the value of x.

The partial areas of any bell-shaped continuous distribution, $f(x)$, whose logarithmic differentials at any point are known, may be found by applying the methods of §§ 3—6. In addition to those given for finding the area of the tail, we may also use the method of § 9 where the tail of a Type III curve is fitted to the tail of a continuous distribution. Clearly if $u(x)$ be the Type III curve fitted by this method to the tail $x \geqslant x_0$ of $f(x)$, then

$$f(\xi) = u(\xi) \exp.(a_4 \xi^4/4! + a_5 \xi^5/5! + \ldots),$$

where $\xi = x - x_0$ and a_i the ith logarithmic differential of $f(x)/u(x)$. Moreover,

$$f(\xi) = u(\xi)[1 + a_4 \xi^4/4! + a_5 \xi^5/5! + \ldots].$$

The area $\int_{x_0}^{\infty} f(x)\,dx$ of the tail of $f(x)$ may then be expressed in terms of incomplete Gamma functions.

In conclusion I wish to record my indebtedness to Professor Karl Pearson for suggesting this problem to me and for his helpful advice and criticism. I wish also to thank Miss Kirby for preparing my diagrams.

DIE STATISTIK DER SELTENEN EREIGNISSE*.

VON ROLF LÜDERS.

I. TEIL. THEORIE.

§ 1. *Einleitung und Problemstellung.* Werden n Individuen eines Kollektivs auf das Vorhandensein eines gewissen Merkmals geprüft und ist bei jedem Individuum die Wahrscheinlichkeit für das Vorhandensein dieses Merkmals q, so ist die Wahrscheinlichkeit dafür, dass genau r Individuen das Merkmal besitzen, bekanntlich

$$P_r = \binom{n}{r} q^r (1-q)^{n-r}.$$

Diese Formel, die häufig Newtonsche Formel genannt wird, lässt sich aus dem binomischen Satz leicht herleiten†.

Wird die Anzahl n der Individuen unendlich gross und konvergiert q gegen Null derart, dass $n \cdot q = m$ endlich bleibt, so erhält man die sogenannte Poissonsche Formel

$$P_r = \frac{e^{-m} \cdot m^r}{r!}.$$

Diese Formel stellt zugleich eine Näherungsformel für den Fall dar, dass n sehr gross und q sehr klein ist. Die Grösse m bedeutet dabei den Mittelwert, d. h. die durchschnittliche Zahl der Individuen, die das Merkmal besitzen.

Die Poissonsche Formel lässt sich in zahlreichen Fällen auf die Statistik der seltenen Ereignisse, wie der Unfälle, Krankheiten oder Selbstmorde anwenden. Bei Unfällen ist das Kollektiv die einem Unfall ausgesetzte Bevölkerung, das Merkmal ist das Eintreten eines Unfalls. Stellt man die Zahl der Unfälle etwa für eine Reihe von Monaten fest, so kann man die beobachtete Häufigkeit einer bestimmten Zahl von Unfällen im Monat mit der aus der Poissonschen Formel berechneten Häufigkeit vergleichen.

Die Poissonsche Formel gilt aber nur unter folgenden Voraussetzungen:

(1) Die Wahrscheinlichkeit für den Eintritt des Ereignisses muss in allen beobachteten Monaten dieselbe sein, sie darf nicht von der Zeit abhängen.

(2) Die Wahrscheinlichkeit muss sehr klein sein, da unter dieser Voraussetzung die Poissonsche Formel abgeleitet worden ist.

* Dissertation zur Erlangung der Doktorwürde der Mathematisch-Naturwissenschaftlichen Fakultät der Hamburgischen Universität, 1933.

† Mises, *Vorlesungen aus dem Gebiete der angewandten Mathematik*, I. Bd., Leipzig-Wien, 1931, S. 128.

(3) Die einzelnen Ereignisse müssen unabhängig von einander sein, durch das Eintreten eines Ereignisses darf die Wahrscheinlichkeit nicht geändert werden. Es darf also keine Wahrscheinlichkeitsansteckung bestehen.

Lässt man die letzte Voraussetzung fallen, so kommt man zu der von Eggenberger und Pólya aufgestellten Formel, die im folgenden stets Eggenbergersche Formel genannt wird*. Nach dieser Formel ist die Wahrscheinlichkeit dafür, dass genau r Ereignisse eintreten,

$$P_r = \frac{1}{r!} h(h+d)(h+2d)\dots(h+(r-1)d).(1+d)^{-\frac{h}{d}-r} \qquad (r=1, 2, 3, \dots);$$

und es ist $P_0 = (1+d)^{-\frac{h}{d}}$.

Dabei ist h der Mittelwert, die Grösse d ergibt sich aus der Beziehung

$$h(1+d) = \mu_2, \text{ wo } \mu_2 = \sum_{r=0}^{\infty} (r-m)^2 P_r,$$

das Mittel zweiter Ordnung (Quadrat der mittleren Abweichung oder Streuungsquadrat) ist.

Die Eggenbergersche Formel findet Verwendung in der Statistik der ansteckenden Krankheiten. Ihre Ableitung aus dem Urnenschema erscheint aber, worauf in einem Artikel von H. Pollaczek-Geiringer† hingewiesen wird, sehr gekünstelt. In demselben Artikel gibt nun die Verfasserin eine Formel an, die einen weit allgemeineren Ansatz für die Theorie der seltenen ansteckenden Ereignisse bietet.

Diese Formel, zur Abkürzung als "Formel I" bezeichnet, soll im folgenden von einem speziellen Standpunkt aus hergeleitet werden. Darauf sind einige Spezialfälle zu diskutieren, ausserdem ist der Zusammenhang mit der Eggenbergerschen Formel zu klären. Ferner soll die Frage der Mittelwerte einer kurzen Untersuchung unterzogen werden.

Die Ergebnisse der Theorie sollen im zweiten Teil an Hand einiger Beispiele nachgeprüft werden.

§ 2. *Ableitung der Formel I.* In der Statistik der seltenen Ereignisse ist die Poissonsche Formel häufig deswegen nicht gültig, weil nicht nur Ereignisse beobachtet werden, die einzeln und unabhängig von einander eintreten, sondern auch Ereignisse, die zu je zweien oder dreien eintreten. Das gilt etwa bei Unfällen: Durch *eine* Ursache kann eine Person zu Schaden kommen, es können aber auch zwei, drei oder noch mehr Todesfälle durch *eine* Ursache eintreten.

Es sei h_1 die mittlere Anzahl (der Mittelwert) der einzeln auftretenden Ereignisse, h_2 die mittlere Anzahl der Paare von Ereignissen, h_3 die mittlere Anzahl der Tripel von Ereignissen, allgemein h_n die mittlere Anzahl der n-fachheit von Ereignissen. Dann ist die mittlere Anzahl der Ereignisse überhaupt

$$m = h_1 + 2h_2 + 3h_3 + \dots + nh_n + \dots.$$

* Pólya und Eggenberger, *Zeitschrift f. angew. Math. u. Mech.* III. 279, 1923.
† *Zeitschr. f. angew. Math. u. Mech.* VIII. 292, 1928.

Nun sei die fundamentale Annahme gemacht, dass die Wahrscheinlichkeiten für das Auftreten von 1, 2, 3 bzw. allgemein n zusammengehörigen Ereignissen durch die Poissonsche Formel bestimmt werden und von einander unabhängig sind. Dann ist die Wahrscheinlichkeit dafür, dass genau ν_1 Einzelereignisse vorkommen,

$$P_{\nu_1}^{(1)} = \frac{e^{-h_1} \cdot h_1^{\nu_1}}{\nu_1!},$$

entsprechend die Wahrscheinlichkeit für das Vorkommen von ν_2 Paaren von Ereignissen

$$P_{\nu_2}^{(2)} = \frac{e^{-h_2} \cdot h_2^{\nu_2}}{\nu_2!},$$

und ganz allgemein für das Vorkommen von ν_n n-fachheiten von Ereignissen

$$P_{\nu_n}^{(n)} = \frac{e^{-h_n} \cdot h_n^{\nu_n}}{\nu_n!}.$$

Wegen obiger Voraussetzung der Unabhängigkeit der Verteilungen von einander ist die Wahrscheinlichkeit für das Auftreten von r Ereignissen überhaupt

$$P_r = \sum_{\nu_1+2\nu_2+\ldots+r\nu_r=r} P_{\nu_1}^{(1)} \cdot P_{\nu_2}^{(2)} \ldots P_{\nu_r}^{(r)},$$

$$P_r = e^{-h_1-h_2-\ldots-h_r-\ldots} \sum_{\nu_1+2\nu_2+\ldots+r\nu_r=r} \frac{h_1^{\nu_1} \cdot h_2^{\nu_2} \ldots h_r^{\nu_r}}{\nu_1!\,\nu_2! \ldots \nu_r!}.$$

Aus dieser im folgenden als "Formel I" bezeichneten Formel ergibt sich speziell

$$P_0 = e^{-h_1-h_2-h_3-\ldots},$$

$$P_1 = P_0 \cdot h_1, \quad P_2 = P_0\left(\frac{h_1^2}{2!} + h_2\right), \quad P_3 = P_0\left(\frac{h_1^3}{3!} + h_1 h_2 + h_3\right).$$

§ 3. *Der Mittelwert und die Mittel höherer Ordnung bei der Formel I.* Zunächst ist die erzeugende Funktion der Formel des vorigen Paragraphen zu bestimmen. Diese Funktion erhält man aus

$$f(z) = \sum_{r=0}^{\infty} P_r z^r = \sum_{r=0}^{\infty} z^r e^{-h_1-h_2-\ldots} \sum_{\nu_1+2\nu_2+\ldots=r} \frac{h_1^{\nu_1} h_2^{\nu_2} \ldots}{\nu_1!\,\nu_2! \ldots}$$

$$= e^{-h_1-h_2-\ldots} \sum_{r=0}^{\infty} \sum_{\nu_1+2\nu_2+\ldots=r} \frac{(h_1 z)^{\nu_1} (h_2 z^2)^{\nu_2} \ldots}{\nu_1!\,\nu_2! \ldots},$$

da $z^r = z^{\nu_1+2\nu_2+\ldots}$ ist.

Weiter folgt $$f(z) = e^{-h_1-h_2-\ldots} \sum_{\nu_1=0}^{\infty} \frac{(h_1 z)^{\nu_1}}{\nu_1!} \sum_{\nu_2=0}^{\infty} \frac{(h_2 z^2)^{\nu_2}}{\nu_2!} \ldots.$$

Die erzeugende Funktion der Formel I ist also

$$f(z) = e^{-h_1-h_2-h_3-\ldots} \cdot e^{h_1 z + h_2 z^2 + h_3 z^3 + \ldots}.$$

Zunächst ist $$f(1) = 1, \text{ d. h. } \sum_{r=0}^{\infty} P_r = 1,$$

eine Bedingung, die bekanntlich notwendig erfüllt sein muss.

Um die Mittelwerte zu erhalten, hat man zu differenzieren. Setzt man zur Abkürzung

$$h(z) = -h_1 - h_2 - h_3 - \ldots + h_1 z + h_2 z^2 + h_3 z^3 + \ldots,$$

so erhält man $f(z) = e^{h(z)}$, $f'(z) = h'(z)\, e^{h(z)}$,

$$f''(z) = \{h'(z)^2 + h''(z)\}\, e^{h(z)}, \quad f'''(z) = \{h'(z)^3 + 3h'(z)h''(z) + h'''(z)\}\, e^{h(z)}.$$

Nun ist $h(1) = 0$, $h'(1) = h_1 + 2h_2 + 3h_3 + 4h_4 + \ldots$,

$$h''(1) = 2.1h_2 + 3.2h_3 + 4.3h_4 + \ldots, \quad h'''(1) = 3.2.1h_3 + 4.3.2h_4 + \ldots.$$

Man erhält also zunächst $f'(1) = h_1 + 2h_2 + 3h_3 + \ldots$.

Da nun der Mittelwert $m = \sum_{r=0}^{\infty} r P_r = f'(1)$

ist, so ergibt sich wie im vorigen Paragraphen

$$m = h_1 + 2h_2 + 3h_3 + \ldots.$$

Ferner ist $f''(1) = m^2 + 2.1h_2 + 3.2h_3 + 4.3h_4 + \ldots$.

Andererseits gilt für das Mittel zweiter Ordnung

$$\mu_2 = \sum_{r=0}^{\infty} (r-m)^2 P_r = \sum_{r=0}^{\infty} r(r-1) P_r + m - m^2$$
$$= f''(1) + m - m^2;$$

also wird

$$\mu_2 = h_1 + 2h_2 + 3h_3 + 4h_4 + \ldots$$
$$+ 2.1h_2 + 3.2h_3 + 4.3h_4 + \ldots$$
$$= h_1 + 4h_2 + 9h_3 + 16h_4 + \ldots,$$

oder $\mu_2 = h_1 + 2^2 h_2 + 3^2 h_3 + 4^2 h_4 + \ldots$.

Endlich ist

$$f'''(1) = m^3 + 3m(2.1h_2 + 3.2h_3 + 4.3h_4 + \ldots) + 3.2.1h_3 + 4.3.2h_4 + \ldots.$$

Nun gilt für das Mittel dritter Ordnung

$$\mu_3 = \sum_{r=0}^{\infty} (r-m)^3 P_r = \sum_{r=0}^{\infty} r(r-1)(r-2) P_r + 3(1-m) \sum_{r=0}^{\infty} r(r-1) P_r + m - 3m^2 + 2m^3$$
$$= f'''(1) + 3(1-m) f''(1) + m - 3m^2 + 2m^3;$$

also wird

$$\mu_3 = m^3 + 3mh''(1) + h'''(1) + 3m^2 + 3h''(1) - 3m^3 - 3mh''(1) + m - 3m^2 + 2m^3$$
$$= h'''(1) + 3h''(1) + m$$
$$= h_1 + 2h_2 + 3h_3 + 4h_4 + \ldots$$
$$+ 3.2.1h_2 + 3.3.2h_3 + 3.4.3h_4 + \ldots$$
$$+ 3.2.1h_3 + 4.3.2h_4 + \ldots$$
$$\ldots\ldots\ldots\ldots,$$

d. h. aber $\mu_3 = h_1 + 8h_2 + 27h_3 + 64h_4 + \ldots$

oder $\mu_3 = h_1 + 2^3 h_2 + 3^3 h_3 + 4^3 h_4 + \ldots$.

Für die Berechnung der Parameter stehen also zunächst die folgenden drei Gleichungen zur Verfügung:

$$m = h_1 + 2h_2 + 3h_3 + \ldots,$$
$$\mu_2 = h_1 + 2^2 h_2 + 3^2 h_3 + \ldots,$$
$$\mu_3 = h_1 + 2^3 h_2 + 3^3 h_3 + \ldots.$$

Die Mittel vierter und höherer Ordnung dürften für die praktische Verwendung kaum in Frage kommen, da sie mit zu hohen Fehlern behaftet sind.

§ 4. *Die Eggenbergersche Formel als Spezialfall der Formel I.* Es ist jetzt zu zeigen, dass die Formel von Eggenberger sich als Spezialfall der Formel I ansehen lässt. Sollen die Verteilungen nach diesen Formeln übereinstimmen, so muss die Gleichung

$$e^{-h_1-h_2-h_3-\ldots} = (1+d)^{-\frac{h}{d}},$$

und ausserdem die Gleichung

$$e^{-h_1-h_2-h_3-\ldots} \sum_{\nu_1+2\nu_2+3\nu_3+\ldots=r} \frac{h_1^{\nu_1} h_2^{\nu_2} h_3^{\nu_3} \ldots}{\nu_1!\,\nu_2!\,\nu_3! \ldots} = (1+d)^{-\frac{h}{d}} \cdot \frac{h\,(h+d)\ldots(h+(r-1)\,d)}{r!\,(1+d)^r},$$

für $r = 1, 2, 3, \ldots$ erfüllt sein. Es müssen demnach die folgenden Gleichungen gelten:

$$h_1 = \frac{h}{1+d} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1),$$

$$\frac{h_1^2}{2!} + h_2 = \frac{h\,(h+d)}{2!\,(1+d)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2),$$

$$\frac{h_1^3}{3!} + h_1 h_2 + h_3 = \frac{h\,(h+d)\,(h+2d)}{3!\,(1+d)^3} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3),$$

usw.

Aus (2) erhält man unter Benutzung von (1)

$$\frac{h^2}{2!\,(1+d)^2} + h_2 = \frac{h^2}{2!\,(1+d)^2} + \frac{hd}{2\,(1+d)^2},$$

$$h_2 = \frac{hd}{2\,(1+d)^2}.$$

Analog ergibt sich aus (3) unter Benutzung der Gleichungen (1) und (2)

$$\frac{h^3}{3!\,(1+d)^3} + \frac{h^2 d}{2\,(1+d)^3} + h_3 = \frac{h^3}{3!\,(1+d)^3} + \frac{h^2 d}{2\,(1+d)^3} + \frac{hd^2}{3\,(1+d)^3},$$

$$h_3 = \frac{hd^2}{3\,(1+d)^3}.$$

Setzt man allgemein für $n = 1, 2, 3, \ldots$

$$h_n = \frac{hd^{n-1}}{n\,(1+d)^n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4),$$

so wird aus der erzeugenden Funktion der Formel I (§ 3)

$$f(z) = e^{-h_1-h_2-h_3-\cdots} \cdot e^{h_1 z + h_2 z^2 + h_3 z^3 + \cdots},$$

wegen $$h_1 z + h_2 z^2 + h_3 z^3 + \ldots = \frac{h}{d}\left\{\frac{dz}{1+d} + \frac{d^2 z^2}{2(1+d)^2} + \frac{d^3 z^3}{3(1+d)^3} + \ldots\right\}$$

$$= -\frac{h}{d} ln\left(1 - \frac{dz}{1+d}\right),$$

die Funktion $$f(z) = e^{\frac{h}{d} ln\left(1-\frac{d}{1+d}\right)} \cdot e^{-\frac{h}{d} ln\left(1-\frac{dz}{1+d}\right)}$$

$$= \left(1 - \frac{d}{1+d}\right)^{\frac{h}{d}} \cdot \left(1 - \frac{dz}{1+d}\right)^{-\frac{h}{d}},$$

$$f(z) = (1+d)^{-\frac{h}{d}} \cdot \left(1 - \frac{dz}{1+d}\right)^{-\frac{h}{d}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5).$$

Andererseits ist bei der Eggenbergerschen Verteilung, wie Eggenberger nachgewiesen hat*, die erzeugende Funktion

$$\bar{f}(z) = \sum_{r=0}^{\infty} \bar{P}_r z^r = (1+d)^{-\frac{h}{d}}\left\{1 + \sum_{r=1}^{\infty} \frac{h(h+d)\ldots(h+(r-1)d)}{r!\,(1+d)^r} z^r\right\}$$

$$= (1+d)^{-\frac{h}{d}}\left\{1 + \sum_{r=1}^{\infty} \frac{1}{r!}\left(-\frac{h}{d}\right)\left(-\frac{h}{d}-1\right)\ldots\left(-\frac{h}{d}-r+1\right)\cdot\left(-\frac{dz}{1+d}\right)^r\right\}$$

$$= (1+d)^{-\frac{h}{d}}\left\{1 + \sum_{r=1}^{\infty} \left(-\frac{h}{d}\right)_r \left(-\frac{dz}{1+d}\right)^r\right\},$$

$$\bar{f}(z) = (1+d)^{-\frac{h}{d}}\left(1 - \frac{dz}{1+d}\right)^{-\frac{h}{d}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6).$$

Hieraus folgt wegen (5), dass die erzeugenden Funktionen identisch sind, d. h. dass $f(z) = \bar{f}(z)$,

oder $$\sum_{r=0}^{\infty} P_r z^r = \sum_{r=0}^{\infty} \bar{P}_r z^r \text{ ist.}$$

Setzt man $z = 0$, so folgt $$P_0 = \bar{P}_0.$$

Ferner ist für $k = 1, 2, 3, \ldots$

$$\frac{d^k}{dz^k}\left(\sum_{r=0}^{\infty} P_r z^r\right) = \frac{d^k}{dz^k}\left(\sum_{r=0}^{\infty} \bar{P}_r z^r\right);$$

setzt man wieder $z = 0$, so erhält man die Gleichungen

$$P_k = \bar{P}_k \qquad (k = 1, 2, 3, \ldots).$$

Sind also die erzeugenden Funktionen zweier Verteilungen identisch, so stimmen beide Verteilungen überein. Damit ist bewiesen, dass die Eggenbergersche Formel ein Spezialfall der Formel I ist. Man erhält die Formel von Eggenberger, wenn man bei der Formel I

$$h_n = \frac{h d^{n-1}}{n(1+d)^n} \qquad (n = 1, 2, 3, \ldots)$$

setzt.

* *Die Wahrscheinlichkeitsansteckung*, Bern, 1924.

§ 5. *Spezialfälle der Formel I.* Die Formel I ist zunächst nicht zu verwenden, da unendlich viele Parameter $h_1, h_2, h_3, \ldots$ bestimmt werden müssen. Es lassen sich aber spezielle Annahmen über diese Grössen machen, sodass nur eine endliche Anzahl von Parametern zu berechnen ist.

Die Poissonsche Formel erhält man, wenn man

$$h_2 = h_3 = h_4 = \ldots = 0$$

setzt; denn dann ist $h_1 = m$

und $$P_r = \frac{e^{-m} \cdot m^r}{r!} \qquad (r = 0, 1, 2, \ldots).$$

Man kann auch annehmen, dass h_1 und h_2 von Null verschieden, dagegen $h_3 = h_4 = \ldots = 0$ ist, d. h. dass die Ereignisse entweder isoliert oder zu Paaren auftreten können, jedoch nicht mehr zu dreien, vieren usw. Unter dieser Voraussetzung ergibt sich die Formel I*a*:

$$P_r = e^{-h_1-h_2} \sum_{\nu_1+2\nu_2=r} \frac{h_1^{\nu_1} h_2^{\nu_2}}{\nu_1!\,\nu_2!}.$$

Dabei ist $$h_1 + 2h_2 = m, \quad h_1 + 4h_2 = \mu_2,$$

also $$h_1 = 2m - \mu_2, \quad h_2 = \frac{\mu_2 - m}{2}.$$

Die Parameter lassen sich auf eine recht einfache Weise berechnen.

Analog erhält man unter der Voraussetzung, dass höchstens drei zusammengehörige Ereignisse beobachtet werden,

$$h_4 = h_5 = h_6 = \ldots = 0.$$

Die entsprechende Formel I*b* lautet:

$$P_r = e^{-h_1-h_2-h_3} \sum_{\nu_1+2\nu_2+3\nu_3=r} \frac{h_1^{\nu_1} h_2^{\nu_2} h_3^{\nu_3}}{\nu_1!\,\nu_2!\,\nu_3!}.$$

Dabei ist

$$\begin{aligned} h_1 + 2h_2 + 3h_3 &= m, \\ h_1 + 4h_2 + 9h_3 &= \mu_2, \\ h_1 + 8h_2 + 27h_3 &= \mu_3, \end{aligned}$$

woraus folgt

$$\begin{aligned} h_1 &= 3m - \frac{5}{2}\mu_2 + \frac{\mu_3}{2}, \\ h_2 &= -\frac{3}{2}m + 2\mu_2 - \frac{\mu_3}{2}, \\ h_3 &= \frac{m}{3} - \frac{\mu_2}{2} + \frac{\mu_3}{6}. \end{aligned}$$

§ 6. *Die Formel II, eine Erweiterung der Eggenbergerschen Formel.* Man erhält, wie in § 4 gezeigt worden ist, die Eggenbergersche Formel aus Formel I, wenn man die $h_1, h_2, h_3, \ldots$ in der folgenden Progression abnehmen lässt:

$$h_1, \quad h_2 = \frac{h_1 q}{2}, \quad h_3 = \frac{h_1 q^2}{3}, \quad h_4 = \frac{h_1 q^3}{4}, \quad \ldots, \quad h_n = \frac{h_1 q^{n-1}}{n}, \quad \ldots,$$

wo $q = \frac{h}{1+d}$ ist. Nun hat die Formel von Eggenberger aber nur zwei Parameter (h und d). Um eine Formel mit drei Parametern zu bekommen, liegt es nahe, die $h_1, h_2, h_3, \ldots$ in der Weise zu spezialisieren, dass man nach Festsetzung von h_1 und h_2 erst die $h_3, h_4, h_5, \ldots$ in entsprechender Progression abnehmen lässt:

$$h_1,\quad h_2 = \frac{p}{2},\quad h_3 = \frac{pq}{3},\quad h_4 = \frac{pq^2}{4},\quad \ldots,\quad h_n = \frac{pq^{n-2}}{n},\quad \ldots.$$

Nun ist $$m = h_1 + 2h_2 + 3h_3 + 4h_4 + \ldots,$$

also erhält man $$m = h_1 + p(1 + q + q^2 + \ldots) = h_1 + \frac{p}{1-q}.$$

Ferner ist $$\mu_2 = h_1 + 4h_2 + 9h_3 + 16h_4 + \ldots$$
$$= h_1 + p(2 + 3q + 4q^2 + 5q^3 + \ldots).$$

Es ist jedoch $$2 + 3q + 4q^2 + \ldots = \frac{1}{1-q} + \frac{1}{(1-q)^2}.$$

Daher ist $$\mu_2 = h_1 + p\left(\frac{1}{1-q} + \frac{1}{(1-q)^2}\right).$$

Endlich gilt $$\mu_3 = h_1 + 8h_2 + 27h_3 + 64h_4 + 125h_5 + \ldots$$
$$= h_1 + p(4 + 9q + 16q^2 + 25q^3 + \ldots).$$

Nun ist aber $$4 + 9q + 16q^2 + 25q^3 + \ldots = \frac{1}{1-q} + \frac{1}{(1-q)^2} + \frac{1}{(1-q)^3}.$$

Man hat also das Gleichungssystem:

$$m = h_1 + \frac{p}{1-q} \quad \ldots\ldots\ldots\

auszuwerten. Nun ist

$$h_1 + h_2 + h_3 + \ldots = h_1 + \frac{p}{2} + \frac{pq}{3} + \frac{pq^2}{4} + \ldots$$

$$= h_1 - \frac{p}{q} + \frac{p}{q^2}\left(q + \frac{q^2}{2} + \frac{q^3}{3} + \frac{q^4}{4} + \ldots\right)$$

$$= h_1 - \frac{p}{q} - \frac{p}{q^2} ln\,(1-q).$$

Also wird

$$P_0 = e^{-h_1 + \frac{p}{q} + \frac{p}{q^2} ln\,(1-q)}$$

$$= (1-q)^{\frac{p}{q^2}} \cdot e^{-h_1 + \frac{p}{q}},$$

und man erhält

$$P_r = P_0 \sum_{\nu_1 + 2\nu_2 + 3\nu_3 + \ldots = r} \frac{h_1^{\nu_1} h_2^{\nu_2} h_3^{\nu_3} \ldots}{\nu_1!\,\nu_2!\,\nu_3! \ldots},$$

wobei man für die h_1, h_2, h_3,... die zu Beginn des Paragraphen angegebenen Werte einzusetzen hat.

Die hergeleitete Formel (im folgenden als Formel II bezeichnet) hat gegenüber der Eggenbergerschen Formel den Vorzug, dass die P_r von *drei* Parametern, nämlich h, p, q, abhängen. Sie hat allerdings den Nachteil, dass zur Berechnung der Parameter auch das Mittel dritter Ordnung μ_3 gebraucht wird, das bekanntlich mit einem ziemlich hohen wahrscheinlichen Fehler behaftet ist.

Für

$$p = h_1 q$$

erhält man wieder die Eggenbergersche Formel; es ist dann

$$h_1 = \frac{h}{1+d},$$

$$q = \frac{d}{1+d},$$

also

$$h_n = \frac{p \cdot q^{n-2}}{n} = \frac{h d^{n-1}}{n\,(1+d)^n} \qquad (n = 1, 2, 3, \ldots).$$

Die Formel II stellt daher zugleich eine Verallgemeinerung der Formel von Eggenberger dar.

§ 7. *Modifizierte Mittelwerte.* Die betrachteten Formeln, vor allem die Formel II, haben den Nachteil, dass zur Berechnung der Parameter neben dem Mittelwert m noch die höheren Mittel μ_2 und μ_3 benötigt werden. Nun ist bekanntlich der mittlere Fehler von μ_2

$$\epsilon_{\mu_2} = \sqrt{\frac{2m^2 + m}{N}},$$

der von μ_3

$$\epsilon_{\mu_3} = \sqrt{\frac{15m^3 + 24m^2 + m}{N}},$$

wo N die Zahl der beobachteten Zeiteinheiten ist. (Diese Ausdrücke gelten, streng genommen, nur im Falle der Poissonschen Verteilung; sie sind aber auch bei anderen Verteilungen anwendbar, wenn es nur auf die Grössenordnung der Fehler ankommt.)

Für grössere Werte von m sind die obigen mittleren Fehler offenbar recht hoch. Die Parameter der Formeln, die aus den Mitteln berechnet werden, sind dann so ungenau, dass auf diesem Wege die theoretische Verteilung nicht gefunden werden kann.

Der grosse Fehler der höheren Mittel kommt daher, dass in der Formel

$$\mu_k = \sum_{r=0}^{\infty} (r-m)^k P_r,$$

schon von $k=3$ an die Faktoren $(r-m)^k$ zu gross und deshalb die Summanden, mithin auch die Summe, zu gross werden. Dasselbe gilt von den Ausdrücken $f'(1)$, $f''(1)$ und $f'''(1)$ des § 3. Aus diesem Grunde liegt es nahe, jeden Summanden mit einem Faktor zu multiplizieren, der für grosse r sehr klein wird.

Es soll die folgende Funktion eingeführt werden:

$$\phi(z) = \sum_{r=0}^{\infty} P_r x^r z^r \quad \ldots\ldots\ldots\

ist. Für $z = 1$ erhält man $\phi = e^h$,

$$\phi' = \phi \,.\, h' \quad \ldots\ldots\ldots\

Die drei Gleichungen für h_1, p und q lauten daher

$$h_1 + \frac{px_1}{1-qx_1} = d_1 \quad \ldots\ldots(3),$$

$$h_1 + \frac{px_2}{1-qx_2} = d_2 \quad \ldots\ldots(4),$$

$$h_1 + \frac{px_3}{1-qx_3} = d_3 \quad \ldots\ldots(5).$$

Aus (3) und (4) folgt

$$\frac{px_1}{1-qx_1} - \frac{px_2}{1-qx_2} = d_1 - d_2,$$

d. h.
$$p(x_1 - x_2) = (d_1 - d_2)(1-qx_1)(1-qx_2).$$

Analog ergibt sich aus (4) und (5)

$$p(x_2 - x_3) = (d_2 - d_3)(1-qx_2)(1-qx_3).$$

Durch Division erhält man

$$\frac{1-qx_1}{1-qx_3} = \frac{x_1 - x_2}{x_2 - x_3} \cdot \frac{d_2 - d_3}{d_1 - d_2} \quad \ldots\ldots(6).$$

Aus dieser Gleichung lässt sich q leicht berechnen, nachdem die Werte der d_1, d_2, d_3 entsprechend der Gleichung (2) festgestellt worden sind.

II. Teil. Anwendungen.

§ 9. *Vorbemerkungen zu den Beispielen.* Die im ersten Teil besprochenen Formeln werden im folgenden Teil auf verschiedene Gebiete der Statistik der seltenen Ereignisse angewandt.

Zur Berechnung der Parameter der Formeln werden benutzt:

der Mittelwert $$m = \sum_{r=0}^{\infty} \frac{rP_r}{N} \quad \ldots\ldots(1),$$

das Mittel zweiter Ordnung $$\mu_2 = \sum_{r=0}^{\infty} \frac{(r-m)^2 P_r}{N} \quad \ldots\ldots(2),$$

und das Mittel dritter Ordnung $$\mu_3 = \sum_{r=0}^{\infty} \frac{(r-m)^3 P_r}{N} \quad \ldots\ldots(3).$$

Dabei ist N die Anzahl der Zeitintervalle, über die sich die Beobachtung ausdehnt, und P_r die Anzahl der Zeitintervalle, in denen r Fälle beobachtet worden sind, sodass

$$\sum_{r=0}^{\infty} P_r = N \text{ ist.}$$

Zu bemerken ist hier, dass der wahrscheinlichste Wert des Mittels zweiter Ordnung

$$\sum_{r=0}^{\infty} \frac{(r-m)^2 P_r}{N-1}$$

ist. Gleichwohl ist das Mittel zweiter Ordnung stets nach Gleichung (2) berechnet worden. Der Fehler kann für grosse N vernachlässigt werden. Bei dieser Methode vereinfacht sich die Rechnung durch folgende Gleichungen:

$$\mu_2 = \sum_{r=0}^{\infty} \frac{(r-m+\delta)^2 P_r}{N} - \delta^2,$$

$$\mu_3 = \sum_{r=0}^{\infty} \frac{(r-m+\delta)^3 P_r}{N} - 3\delta\mu_2 - \delta^3.$$

Hier ist δ eine Zahl zwischen 0 und 1, die so gewählt ist, dass $m-\delta$ eine ganze Zahl wird ($m-\delta$ ist ein "vorläufiger Mittelwert," nämlich die dem Mittelwert m nächstkommende ganze Zahl).

Was die Genauigkeit anbelangt, so ist der mittlere Fehler des Mittelwertes m

$$\epsilon_m = \sqrt{\frac{\mu_2}{N}},$$

der des Mittels zweiter Ordnung

$$\epsilon_{\mu_2} = \sqrt{\frac{2m^2+m}{N}},$$

und der des Mittels dritter Ordnung

$$\epsilon_{\mu_3} = \sqrt{\frac{15m^3+24m^2+m}{N}}.$$

(Betreffs der beiden letzten Ausdrücke gilt das in § 7 Gesagte.)

Um die Güte der Anpassung der theoretischen Häufigkeiten an die beobachteten zu prüfen, ist das Pearsonsche χ^2-Kriterium* benutzt worden. Aus der Grösse

$$\chi^2 = \sum_r \frac{(\bar{P}_r - P_r)^2}{\bar{P}_r},$$

wo die P_r die beobachteten, die $\bar{P}_r$ die theoretischen Häufigkeiten sind, erhält man mit Hilfe der Pearsonschen Tabelle die Wahrscheinlichkeit P dafür, dass die Abweichungen von den theoretischen Werten ebenso gross oder grösser als beobachtet sind. (Die Häufigkeiten für grössere r sind bei der Anwendung des Kriteriums zusammenzufassen.) Je näher P an 1 liegt, um so besser ist die Anpassung der Theorie an die Beobachtung.

Grundsätzlich ist noch zu bemerken, dass eine Formel sich um so besser der Beobachtung anpasst, je mehr Parameter sie besitzt. Das gilt insbesondere bei Verteilungen mit sehr kleinen Mittelwerten. In solchen Fällen ist die gute Anpassung einer Formel mit drei Parametern noch kein Beweis dafür, dass das Schema der Formel (etwa die Ansteckung) den tatsächlichen Verhältnissen entspricht.

Die in den folgenden Tabellen angeführten theoretischen Häufigkeiten sind im allgemeinen auf eine Stelle nach dem Komma abgerundet. Dadurch ist es zu

* Pearson, *Tables for Statisticians and Biometricians*, Cambridge, 1914, Table XII.

erklären, dass die Summe dieser Zahlen in einigen Fällen von der Summe N der beobachteten Häufigkeiten um eine oder zwei Einheiten der ersten Dezimale abweicht.

§ 10. *Todesfälle durch Eisen- und Strassenbahn im Saargebiet.* Das Statistische Amt des Saargebietes veröffentlicht die monatliche Anzahl der durch Eisen- und Strassenbahn verursachten Todesfälle*. Es sollen die Monate Januar 1925 bis Dezember 1929, also insgesamt 60 Monate, und zwar nur die Todesfälle von *männlichen* Personen untersucht werden. Tabelle I gibt in Spalte II die Anzahl der Monate an, in denen die in Spalte I angegebene Zahl von Todesfällen vorgekommen ist, also die beobachtete Häufigkeit der Todesfälle. Der Mittelwert beträgt $m = 1{\cdot}383$, d. h. es sind monatlich im Durchschnitt 1·383 Todesfälle eingetreten.

Nach der Formel von Poisson ist die Wahrscheinlichkeit dafür, dass genau r Todesfälle während eines Monats vorkommen,

$$\frac{e^{-m} \cdot m^r}{r!},$$

also die theoretische Häufigkeit, da 60 Monate beobachtet worden sind,

$$P_r = 60 \frac{e^{-m} \cdot m^r}{r!} \qquad (r = 0, 1, 2, \ldots).$$

Spalte III der Tabelle I enthält die auf diese Weise berechneten Häufigkeitszahlen.

Um die Güte der Anpassung an die Beobachtung zu prüfen, hat man nach dem Pearsonschen χ^2-Kriterium die Grösse

$$\chi^2 = \sum_{r=0}^{3} \frac{(P_r - \bar{P}_r)^2}{\bar{P}_r} = 3{\cdot}72$$

zu berechnen. Dabei sind die Häufigkeiten von $r = 4$ ab zusammengefasst; die Anzahl der Summanden ist demnach $n' = 5$. Auf Grund der Pearsonschen Tabelle besteht dann die Wahrscheinlichkeit $P = 0{\cdot}45$ dafür, dass die Abweichungen von der theoretischen Verteilung mindestens ebenso gross wie beobachtet sind.

Bei der Poissonschen Formel müsste

$$m = \mu_2 = \mu_3$$

sein, hier ist aber $m = 1{\cdot}383$, $\mu_2 = 2{\cdot}036$, $\mu_3 = 3{\cdot}585$.

Unter Berücksichtigung der mittleren Fehler

$$\epsilon_m = 0{\cdot}18, \quad \epsilon_{\mu_2} = 0{\cdot}29, \quad \epsilon_{\mu_3} = 1{\cdot}2,$$

dürften die Differenzen zwischen den Mitteln zu gross sein, als dass man sie noch als zufällig ansehen kann.

Was nun die Eggenbergersche Formel angeht, so ist

$$h = m = 1{\cdot}383,$$

$$1 + d = \frac{\mu_2}{m} = 1{\cdot}472, \text{ also } d = 0{\cdot}472.$$

* *Bericht des Statistischen Amtes des Saargebietes*, Saarbrücken; 5. Heft (1927), bis 9. Heft (1929).

Man erhält $$P_0 = 60\,(1+d)^{-\frac{h}{d}} = 19{\cdot}31$$

und die weiteren in Spalte IV angegebenen Häufigkeitszahlen. Die Übereinstimmung mit der Beobachtung ist hier besser als bei der Poissonschen Formel. Denn nach dem Pearsonschen χ^2-Kriterium ist wegen $\chi^2 = 0{\cdot}97$ die Wahrscheinlichkeit dafür, dass die Abweichungen mindestens ebenso gross wie beobachtet sind, $P = 0{\cdot}91$.

In dem vorliegenden Beispiel besteht Ansteckung in dem Sinne, dass durch *einen* Unfall auch zwei oder mehr Personen in Mitleidenschaft gezogen werden können. Nimmt man eine maximale Zahl von drei Toten bei einem Unfall an, so hat man die Formel I *b* anzuwenden. Dann ist

$$h_1 = 3m - \frac{5}{2}\mu_2 + \frac{\mu_3}{2} = 0{\cdot}852,$$

$$h_2 = -\frac{3}{2}m + 2\mu_2 - \frac{\mu_3}{2} = 0{\cdot}205,$$

$$h_3 = \frac{m}{3} - \frac{\mu_2}{2} + \frac{\mu_3}{6} = 0{\cdot}041.$$

Man erhält $$P_0 = 60\,.\,e^{-h_1-h_2-h_3} = 20{\cdot}01$$

und die weiteren Zahlen der Spalte V. Diese Werte stimmen ausgezeichnet mit den empirischen überein. Auf Grund des Pearsonschen χ^2-Kriteriums ist $\chi^2 = 0{\cdot}65$ und $P = 0{\cdot}96$.

Zur Kontrolle sei noch die Formel II benutzt. Nach dieser Formel ist

$$1 - q = 2\,.\,\frac{\mu_2 - m}{\mu_3 - \mu_2} = 0{\cdot}843, \quad q = 0{\cdot}157;$$

ferner ist $p = (\mu_2 - m)(1-q)^2 = 0{\cdot}464$.

Es wird $$h_1 = m - \frac{p}{1-q} = 0{\cdot}833,$$

$$h_2 = \frac{p}{2} = 0{\cdot}232,$$

$$h_3 = \frac{pq}{3} = 0{\cdot}024,$$

$$h_4 = \frac{pq^2}{4} = 0{\cdot}003 \text{ usw.}$$

Bedenkt man, dass insbesondere μ_3 mit einem recht hohen mittleren Fehler behaftet ist, so sind die Abweichungen der obigen Werte der h_1, h_2, h_3 von den nach Formel I *b* berechneten als gering anzusehen. h_4 ist fast gleich Null, womit die oben gemachte Annahme über die Höchstzahl der Toten bei einem Unfall sich als berechtigt erwiesen hat. Die aus dieser Formel zu berechnende Verteilung würde sich daher nur unwesentlich von der Verteilung in Spalte V unterscheiden. Nach Formel II verlaufen von allen tödlichen Unfällen durch Eisen- oder Strassenbahn rd. 76 % mit einem Todesfall, 21 % mit zwei Todesfällen und 2 % mit drei Todesfällen. Eine Nachprüfung dieser Daten durch die Beobachtung ist auf Grund des

zitierten Berichtes nicht möglich, da Angaben über die Zahl der durch die gleiche Ursache Verunglückten fehlen.

Die Eggenbergersche Formel stimmt hier nicht so gut mit der Beobachtung überein wie die Formel I *b*. Das kommt daher, dass nach der Eggenbergerschen Formel

$$h_1 = \frac{h}{1+d} = 0{\cdot}940,$$

$$h_2 = \frac{1}{2}\,\frac{d}{1+d}\cdot h_1 = 0{\cdot}150,$$

$$h_3 = \frac{1}{3}\left(\frac{d}{1+d}\right)^2 h_1 = 0{\cdot}032,$$

$$h_4 = \frac{1}{4}\left(\frac{d}{1+d}\right)^3 h_1 = 0{\cdot}008$$

ist. Man sieht, dass diese Zahlen von den entsprechenden der Formel I *b* oder II abweichen, dass hier insbesondere h_1 zu gross und h_2 zu klein ist.

Die gute Übereinstimmung zwischen Theorie und Beobachtung ist hier vor allem dadurch begründet, dass das Material homogen ist. Es beträgt nämlich

im Jahre	die Zahl der Todesfälle
1925	12
1926	19
1927	15
1928	18
1929	19

Die Schwankungen von Jahr zu Jahr sind also nicht zu gross.

TABELLE I.

Verteilung der männlichen Todesfälle durch Eisen- und Strassenbahn im Saargebiet.

I. Anzahl der Fälle	II. Beobacht. Verteil.	III. Verteil. nach Poisson	IV. Verteil. nach Eggenberger	V. Verteil. nach Formel I *b*
0	20	15·1	19·3	20·0
1	17	20·8	18·1	17·1
2	11	14·4	11·4	11·4
3	8	6·6	6·0	6·4
4	2	2·3	2·9	3·0
5	—	0·6	1·3	1·3
6	2	0·1	0·5	0·5
7 und mehr	—	0·0	0·3	0·3

§ 11. *Verteilung der weiblichen Diphteriefälle in Böhmen* 1921 *bis* 1929. In einem kürzlich erschienenen Artikel der *Aktuárské Vědy* gibt M. Vacek bei der Untersuchung der Eggenbergerschen Formel als Beispiel u. a. die weiblichen *Diphterie*-Fälle in Böhmen für die 108 Monate der Jahre 1921 bis 1929 an*.

Spalte II der folgenden Tabelle II zeigt die beobachtete, Spalte III die von Vacek auf Grund der Eggenbergerschen Formel berechnete Häufigkeit. Dabei sind der Mittelwert

$$m = h = 15{\cdot}8796$$

und die Grösse

$$d = 13{\cdot}6274$$

zugrundegelegt worden. Was die Übereinstimmung mit der beobachteten Verteilung angeht, so ist

$$\chi^2 = 24{\cdot}4$$

also

$$P = 0{\cdot}08,$$

d.h. die Anpassung der Eggenbergerschen Formel ist hier sehr schlecht.

Das Eggenbergersche Schema ist nicht geeignet, entweder weil das Mittel zweiter Ordnung, aus dem d berechnet worden ist, mit einem zu hohen Fehler behaftet ist, oder weil die speziellen Voraussetzungen, die über die Art der Ansteckung zu machen sind, hier nicht zutreffen. Um beide Fehler möglichst auszuschalten, soll die Formel II benutzt werden, wobei die Parameter durch die Methode der modifizierten Mittelwerte berechnet werden sollen. Dadurch wird der hohe Fehler durch das Mittel dritter Ordnung vermieden.

Setzt man

$$x_1 = 1,$$
$$\log x_2 = -0{\cdot}1\,\dagger, \quad \text{d.h.} \quad x_2 = 0{\cdot}7943,$$
$$\log x_3 = -0{\cdot}2, \quad \text{d.h.} \quad x_3 = 0{\cdot}6310,$$

so erhält man entsprechend der Gleichung (2) des § 8

$$d_1 = 15{\cdot}880, \quad d_2 = 6{\cdot}401, \quad d_3 = 5{\cdot}254.$$

Aus der Gleichung (6) des § 8 wird hier

$$\frac{1-q}{1-0{\cdot}631\,q} = 0{\cdot}1524, \text{ also } q = 0{\cdot}9378, \quad p = 0{\cdot}7315,$$

$$h_1 = 4{\cdot}124, \quad h_2 = 0{\cdot}366, \quad h_3 = 0{\cdot}229, \quad h_4 = 0{\cdot}161 \text{ usw.}$$

Die h_i nehmen nur sehr langsam ab, was auf eine starke Ansteckung hindeutet. Man erhält

$$P_0 = 108\,(1-q)^{\frac{p}{q^2}} e^{-h+\frac{p}{q}} = 0{\cdot}378,$$

und die weiteren in Spalte IV angegebenen Zahlen.

Das Pearsonsche χ^2-Kriterium liefert hier wegen $\chi^2 = 8{\cdot}9$

$$P = 0{\cdot}9,$$

* "Sur la loi de Pólya régissant les faits corrélatifs," *Akt. Vědy*, III. 18, 49, 1932.

† x_2 und x_3 sind so gewählt worden, dass ihre Logarithmen auf volle Zehntel lauten, dass sich also ihre Potenzen leicht berechnen lassen.

d. h. die Anpassung an die Beobachtung ist als befriedigend anzusehen. Gleichwohl fällt auf, dass auch hier die theoretische maximale Häufigkeit hinter der beobachteten zurückbleibt.

Bei ansteckenden Krankheiten scheint die Verwendung von drei Parametern unter Umgehung der höheren Mittel zu brauchbaren Resultaten zu führen. Das gilt vor allem für Verteilungen mit einem grossen Mittelwert wie in dem vorliegenden Fall, wo $m = 15{\cdot}8796$ ist.

TABELLE II.

Verteilung der weiblichen Diphteriefälle in Böhmen 1921 *bis* 1929.

I. Anzahl der Fälle	II. Beobacht. Verteil.	III. Verteil. nach Eggenberger	IV. Verteil. nach Formel II
0	1	4·7	0·4
1	1	5·1	1·6
2	3	**5·2**	3·4
3	4	5·1	5·1
4	7	5·0	6·2
5	6	4·8	**6·5**
6	9	4·6	6·4
7	**10**	4·4	5·9
8	7	4·1	5·5
9	5	3·9	5·0
10	3	3·7	4·5
11	4	3·5	4·2
12	1	3·3	3·8
13	3	3·1	3·5
14	3	3·0	3·2
15	2	2·8	3·0
16 und mehr	39	41·8	39·8

§ 12. *Verteilung der Blutkörperchen.* In einem 1907 erschienenen Artikel der *Biometrika** wird die Poissonsche Formel benutzt, um die Verteilung der Blutkörperchen auf die 400 Quadrate eines Hämacytometers zu untersuchen. Falls die Blutkörperchen unabhängig von einander auf die Fläche verteilt sind, ist auch hier die Poissonsche Formel gültig.

Der Verfasser des erwähnten Artikels hat vier verschiedene Konzentrationen untersucht. In der ersten Konzentration ist die mittlere Anzahl der Blutkörperchen pro Quadrat

$$m = 0{\cdot}6825.$$

Tabelle III enthält in Spalte II die Zahl der Quadrate, in denen die in Spalte I angegebene Anzahl von Blutkörperchen beobachtet worden ist. "Student" hat die Zahlen der Poissonschen Formel berechnet (Spalte III).

Die höheren Mittel werden angegeben als

$$\mu_2 = 0{\cdot}8117, \quad \mu_3 = 1{\cdot}0876.$$

* "Student," "On the Error of Counting with a Haemacytometer," *Biometrika*, Vol. v. p. 351, 1907.

Da die mittleren Fehler

$$\epsilon_m = 0{\cdot}045, \quad \epsilon_{\mu_2} = 0{\cdot}064, \quad \epsilon_{\mu_3} = 0{\cdot}20$$

sind, kann man die Abweichungen der höheren Mittel von m nicht als zufällig ansehen. Aus diesem Grunde sind die Differenzen zwischen Theorie und Beobachtung recht gross. Der Verfasser gibt nämlich an

$$\chi^2 = 9{\cdot}92 \quad \text{und} \quad P = 0{\cdot}04,$$

wobei die Häufigkeiten für 4 und mehr Blutkörperchen zusammengefasst sind.

Zur Erklärung wird angenommen, dass die Blutkörperchen das Bestreben haben, zusammenzuhaften ("...there is also probably a tendency to stick together in groups which was not altogether abolished even by vigorous shaking"). Wenn diese Annahme richtig ist, muss Formel II mit der Beobachtung besser übereinstimmen. Man erhält

$$1 - q = 0{\cdot}9366; \quad q = 0{\cdot}0634, \quad p = 0{\cdot}113;$$

$$h_1 = 0{\cdot}562, \quad h_2 = 0{\cdot}057, \quad h_3 = 0{\cdot}0024, \quad h_4 = 0{\cdot}0001 \text{ usw.}$$

Die Wahrscheinlichkeit für Gruppen von 4 oder mehr Blutkörperchen ist also verschwindend klein. Spalte IV der Tabelle erhält die nach der Formel II berechneten Werte, die hier auf ganze Zahlen abgerundet sind. Was die Güte der Anpassung angeht, so ist

$$\chi^2 = 3{\cdot}33, \quad \text{d. h.} \quad P = 0{\cdot}51.$$

Diese weit bessere Anpassung an die Erfahrung spricht für die erwähnte Gruppenbildung.

Formel I *b* liefert die Konstanten

$$h_1 = 0{\cdot}562, \quad h_2 = 0{\cdot}056, \quad h_3 = 0{\cdot}0029,$$

die fast identisch mit denen der Formel II sind.

Die hier geschilderten Verhältnisse treffen nicht für alle untersuchten Konzentrationen zu. In einigen Fällen passt sich nämlich die Poissonsche Formel der Beobachtung gut an.

TABELLE III.

Verteilung der Blutkörperchen auf 400 *Quadrate eines Hämacytometers.*

I. Zahl der Blutkörperchen pro Quadrat	II. Beobacht. Verteil.	III. Verteil. nach Poisson	IV. Verteil. nach Formel II
0	213	202	215
1	128	138	121
2	37	47	46
3	18	11	14
4	3	2	3
5 und mehr	1	0	1

§ 13. *Zahlen der Blumenkronblätter von Ranunculus bulbosus.* Zur Illustration der Poisson-Charlierschen *B*-Kurve sind von Riebesell* die Zahlen der Blumenkronblätter von *Ranunculus bulbosus* benutzt worden. Die mittlere Anzahl der Blumenkronblätter für die 222 untersuchten Exemplare ist 5·631, das Streuungsquadrat

$$\mu_2 = 0{\cdot}918.$$

Spalte II der Tabelle IV zeigt die beobachtete Häufigkeit der in Spalte I angegebenen Zahlen x der Blumenkronblätter.

In dem zitierten Artikel wird zunächst der Wert $x = 5$ als "Nullpunkt" der Verteilung gewählt, d. h. es ist

$$m = 0{\cdot}631.$$

Dann werden aus der Charlierschen Formel

$$P_x = 222\,\{\psi(r) + c\,.\,\Delta^2\psi(r)\},$$

wo

$$\psi(r) = \frac{e^{-m}\,.\,m^r}{r!},$$

$$c = \frac{\mu_2 - m}{2},$$

$$\Delta^2\psi(r) = \psi(r) - 2\psi(r-1) + \psi(r-2)$$

sind, die in Spalte IV angeführten theoretischen Werte berechnet, wobei $r = x - 5$ ist. Spalte III gibt zum Vergleich die Poissonschen Zahlen an.

Nun ist aber $\mu_3 = 1{\cdot}644.$

Die mittleren Fehler sind

$$\epsilon_m = 0{\cdot}064, \quad \epsilon_{\mu_2} = 0{\cdot}080, \quad \epsilon_{\mu_3} = 0{\cdot}25.$$

Die Abweichungen der Mittel von einander sind also grösser, als die mittleren Fehler zulassen.

Die Formel II liefert die folgenden Zahlen:

$$1 - q = 0{\cdot}7906\,; \quad q = 0{\cdot}2094, \quad p = 0{\cdot}1794,$$

$$h_1 = 0{\cdot}4037, \quad h_2 = 0{\cdot}0897, \quad h_3 = 0{\cdot}0125, \quad h_4 = 0{\cdot}00197 \text{ usw.}$$

Man erhält die theoretische Häufigkeit für das Vorkommen von 5 Blumenkronblättern:

$$P_0 = 222\,(1-q)^{\frac{p}{q^2}}\,.\,e^{-h+\frac{p}{q}} = 133{\cdot}5,$$

und die weiteren in Spalte V angegebenen Zahlen.

Um die Güte der Anpassung zu prüfen, seien die Fälle von 9 und 10 Blumenkronblättern zusammengefasst. Für die Poissonsche Formel ist

$$\chi^2 = 18{\cdot}5,$$

d. h. wegen $n' = 5$, $\quad P = 0{\cdot}001,$

* Riebesell, "Biometrik und Variationsstatistik," *Handbuch der biologischen Arbeitsmethoden*, Abt. v. Teil 2, 1. Hälfte, Wien, 1928, S. 759.

die Formel ist unbrauchbar. Für die Charliersche Verteilung ist

$$\chi^2 = 0{\cdot}99, \text{ d. h. } P = 0{\cdot}91.$$

Für die Formel II ist sogar

$$\chi^2 = 0{\cdot}25 \text{ und } P = 0{\cdot}99;$$

diese Formel passt sich am besten der Erfahrung an, wobei jedoch zu beachten ist, dass drei Parameter zur Berechnung gebraucht worden sind. Es scheint, als ob sich statt ursprünglich *eines* Blumenkronblattes zwei oder mehr entwickeln können (vielleicht durch Zellteilung) und so das Schema der Formel II gültig ist.

TABELLE IV.

Zahlen der Blumenkronblätter von Ranunculus bulbosus.

I. Zahl der Blumenkron-blätter x	II. Beobacht. Häufigkeit*	III. Häufigkeit nach Poisson	IV. Häufigkeit nach Charliers B-Kurve	V. Häufigkeit nach Formel II
5	133	118·2	134·9	133·5
6	55	74·5	51·6	53·9
7	23	23·5	22·5	22·9
8	7	4·9	9·5	8·0
9	2	0·8	2·9	2·6
10	2	0·1	0·6	0·8

Vorstehende Arbeit ist durch eine Anregung von Herrn Prof. Dr. Riebesell, Hamburg, entstanden. Ich bin Herrn Prof. Riebesell darüber hinaus für wichtige Hinweise, insbesondere auf die benutzte Literatur, zu Dank verpflichtet.

* De Vries, *Berichte der deutschen botanischen Gesellschaft*, Jahrg. 12, pp. 203—204, 1894.

ON CERTAIN NON-NORMAL SYMMETRICAL FREQUENCY DISTRIBUTIONS.

By G. H. HANSMANN, M.Sc. (Stellenbosch), B.A. (Oxon.).

Introduction.

In 1889 Professor Karl Pearson introduced a series of curves which enabled in all cases a mathematical curve to be found which had the same first four moments as any given frequency distribution, thus providing for the more usual types of frequency an adequate description in mathematical terms of their distribution as judged by the test of goodness of fit. He reached these curves by integrating

$$\frac{1}{y}\frac{dy}{dx} = \frac{-(x+a)}{c_0 + c_1 x + c_2 x^2 + \dots c_n x^n},$$

retaining only the first three terms of the denominator on the right-hand side.

Beside the need to graduate data, there is another aspect of frequency curves. Mathematically—generally with certain limiting hypotheses—the distribution of certain frequency constants, such as the correlation coefficient, the regression coefficient, the variance, and "Student's" z from normal distributions, has been found. In some of these cases the distributions are exact Pearson curves, in other cases these curves suffice when the sample reaches 25—30 individuals. In a further series of cases it is possible to determine the theoretical moment coefficients, although we cannot find the mathematical form of the curve. Examples are the distribution of the third moment and that of the ordinates of the regression line at a given abscissa, both from a normal population. Both curves are symmetrical and neither curve is exactly described by a Pearson curve with the same first four moments. Furthermore, in the case of the distribution of μ_3, a $\beta_2 = 12{\cdot}5$ has been found. Here the corresponding Pearson curve may give us a very bad fit, because the values* of β_4 and of β_6, as determined from the difference equation, are negative when $\beta_2 > 6$ and $4{\cdot}5 < \beta_2 < 6$ respectively.

We now ask: To what extent shall we better matters by taking into account higher moments? What are the types of curves to which higher moments lead us in such cases? Shall we be able to find a higher order symmetrical Pearson curve when $\beta_2 > 6$ and $4{\cdot}5 < \beta_2 < 6$, for which β_4 and β_6 are respectively positive?

Dr David Heron has already discussed the third order Pearson curves which correspond to the differential equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{-(x+a)}{c_0 + c_1 x + c_2 x^2 + c_3 x^3},$$

* $\beta_4 = \dfrac{5\beta_2^2}{6-\beta_2}$, $\beta_6 = \dfrac{35\beta_2^3}{(6-\beta_2)(9-2\beta_2)}$.

but his manuscript* has remained unpublished because the addition of a single c_3 applied to actually occurring frequencies did not as a rule materially improve the goodness of fit.

The aim of the present paper is to include the following term in c_4, but to consider only the symmetrical curves, for which a, c_1, c_3 may be taken to be zero.

This leads us to the differential equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{-x}{c_0 + c_2x^2 + c_4x^4} \quad \text{.....(1).}$$

It is to the discussion of the integrals of this equation and their applications that this paper is devoted.

I. The Different Fourth Order Pearson Curves.

(*a*) *Discussion of Types.*

Without the loss of generality we can write the differential equation (1) as

$$\frac{1}{y}\frac{dy}{\frac{dx}{\sigma}} = \frac{x/\sigma}{b_0 + b_2\frac{x^2}{\sigma^2} + b_4\frac{x^4}{\sigma^4}},$$

where the constants b_0, b_2 and b_4 are now mere numerics.

Therefore
$$\frac{1}{y}\frac{dy}{dx} = \frac{x}{b_0\sigma^2 + b_2x^2 + b_4\frac{x^4}{\sigma^2}} \quad \text{.....(2).}$$

Multiply both sides by yx^{2s+1} and integrate with respect to x between the limits a_1 and a_2. Divide by N.

Hence
$$\frac{1}{N}\int_{a_1}^{a_2} x^{2s+1}\left(b_0\sigma^2 + b_2x^2 + b_4\frac{x^4}{\sigma^2}\right)dy = \frac{1}{N}\int_{a_1}^{a_2} x^{2s+2}\,y\,dx.$$

Integrate the left-hand side by parts. We find

$$\frac{1}{N}x^{2s+1}\left\{b_0\sigma^2 + b_2x^2 + \frac{b_4}{\sigma^2}x^4\right\}y\Big|_{a_1}^{a_2}$$
$$-\frac{1}{N}\int_{a_1}^{a_2}\left\{(2s+1)\,b_0\sigma^2x^{2s} + (2s+3)\,b_2x^{2s+2} + (2s+5)\frac{b_4}{\sigma^2}x^{2s+4}\right\}y\,dx$$
$$= \frac{1}{N}\int_{a_1}^{a_2} x^{2s+2}y\,dx.$$

Take the origin of x at the mean of the distribution. Now since the distribution is symmetrical the mean and the mode correspond. Hence the limits of integration are from $a_1 = -a$ to $a_2 = +a$ and the integral $\frac{1}{N}\int_{-a}^{+a} x^s\,y\,dx = \mu_s = s$th moment about the mean of the distribution.

* I am very grateful to Prof. K. Pearson for drawing my attention to this manuscript.

Therefore

$$(2s+1)\,\mu_{2s}\,b_0\sigma^2+(2s+3)\,b_2\,\mu_{2s+2}+(2s+5)\,\frac{b_4}{\sigma^2}\,\mu_{2s+4}=c_a-\mu_{2s+2},$$

where
$$c_a=\frac{2a^{2s+1}}{N}\left\{b_0\sigma^2+b_2a^2+\frac{b_4}{\sigma^2}a^4\right\}y_a \quad \text{.....(3).}$$

Now c_a vanishes if

(i) a is zero; but a cannot vanish or there would be no range;

(ii) a^2 is a root of $b_0\sigma^2+b_2a^2+\frac{b_4}{\sigma^2}a^4=0$ and if y_a is finite;

(iii) y_a is zero and if $\left(b_0\sigma^2+b_2a^2+\frac{b_4}{\sigma^2}a^4\right)$ is zero or remains finite.

By means of (ii) and (iii) we are able to fix the limits of integration. Put $\frac{x}{\sigma}=u$ in the equation (2). We obtain

$$\frac{1}{y}\frac{dy}{du}=\frac{u}{b_0+b_2u^2+b_4u^4} \quad \text{.....(4).}$$

Now write $m_1=\frac{1}{2b_4}$, $z=u^2=\frac{x^2}{\sigma^2}$. We obtain

$$\frac{1}{y}\frac{dy}{dz}=\frac{m_1}{\frac{b_0}{b_4}+\frac{b_2}{b_4}z+z^2} \quad \text{.....(5).}$$

The expression $\left(\frac{b_0}{b_4}+\frac{b_2}{b_4}z+z^2\right)$ can assume any of the forms

(1) $(p_1+z)(q_1+z)$, real and unequal factors;

(2) $(b_1+z)^2+c_1^2$, imaginary factors;

(3) $(p_1+z)^2$, equal factors.

Let us integrate the differential equation (5) in these three cases and consider what happens to the expression for c_a.

(1)
$$\frac{1}{y}\frac{du}{dz}=\frac{m_1}{(p_1+z)(q_1+z)}$$
$$=\frac{m_1}{q_1-p_1}\left(\frac{1}{p_1+z}-\frac{1}{q_1+z}\right).$$

Integrating, we find

$$\log y=\text{const.}+\left(\frac{m_1}{q_1-p_1}\right)\log\left(\frac{p_1+z}{q_1+z}\right),$$

therefore
$$y=y_0\left[\frac{p_1+z}{q_1+z}\right]^{\left(\frac{m_1}{q_1-p_1}\right)},$$

where y_0 is the constant of integration.

Notation. I shall use letters with suffix$_1$ to represent the quantities m, q, p, k and k', when they can be positive or negative and unsuffixed letters to represent their absolute values. Write

$$q = |q_1|, \qquad p = |p_1|, \qquad m = |m_1|,$$

$$k_1 = \frac{m_1}{q-p}, \qquad k_1' = \frac{m_1}{q+p},$$

$$k = |k_1|, \qquad k' = |k_1'|.$$

I shall always take $q > p$.

(i) p_1 and q_1 both positive. Hence writing $p = p_1$ and $q = q_1$ we find

$$c_a = \text{const.} \times a^{2s+1}(a^2+p)^{k_1+1}(a^2+q)^{1-k_1}.$$

Hence $c_a = 0$ if $a^2 = -p$, when we have an imaginary range, or if $a^2 = \infty$, provided in addition $(2s+1) + 2(k_1+1) < 2(k_1-1)$, i.e. $2s < -5$. Hence c_a can only be zero for negative moments.

(ii) $p_1 < 0$ and $q_1 > 0$. Write $p = -p_1$ and $q = q_1$. Clearly the equation for y is now

$$y = y_0\left(\frac{p-u^2}{q+u^2}\right)^{k_1'}.$$

Accordingly, $c_a = \text{const.} \times a^{2s+1}(p-a^2)^{k_1'+1}(q+a^2)^{1-k_1'}$ and becomes zero at $a = \pm\sqrt{p}$, provided $k_1' \geqq -1$. If k_1' be < -1, c_a can never be zero.

We get a similar discussion for $p_1 > 0$ and $q_1 < 0$.

(iii) $p_1 < 0$ and $q_1 < 0$. Write $p_1 = -p$ and $q_1 = -q$. The equation for y reduces to

$$y = y_0\left(\frac{q-u^2}{p-u^2}\right)^{k_1}, \text{ where } k_1 = \frac{m_1}{q-p}.$$

We now find $c_a = \text{const.} \times [p-a^2]^{1-k_1}[q-a^2]^{k_1+1}a^{2s+1}$ and accordingly $c_a = 0$ at $a = \pm\sqrt{p}$, provided $k_1 < 1$, but can never be $= 0$ when $k_1 \geqq 1$.

(2) $$\frac{1}{y}\frac{dy}{du} = \frac{2m_1u}{(u^2+b_1)^2+c_1^2}.$$

Integrating with respect to u^2 we obtain

$$\log y = \text{const.} + \frac{m_1}{c_1}\tan^{-1}\frac{b_1+u^2}{c_1},$$

therefore $$y = y_0 e^{\frac{m_1}{c_1}\tan^{-1}\frac{b_1+u^2}{c_1}}.$$

Hence $$c_a = \text{const.} \times a^{2s+1}e^{\frac{m_1}{c_1}\tan^{-1}\frac{b_1+a^2}{c_1}}[(a^2+b_1)^2+c_1^2].$$

Clearly c_a can only be zero if $a = \pm\infty$ and if in addition we have the condition $(2s+5) < 0$ or $s < -\frac{5}{2}$. In other words, for positive moments c_a cannot be zero.

(3) $$\frac{1}{y}\frac{dy}{du} = \frac{2m_1 u}{(p_1 + u^2)^2},$$

therefore $$\log y = \text{const.} - \frac{m_1}{p_1 + u^2},$$

therefore $$y = y_0 e^{-\frac{m_1}{p_1+u^2}}.$$

Accordingly $$c_a = \text{const.} \times a^{2s+1} e^{-\frac{m_1}{p_1+a^2}} (p_1 + a^2)^2.$$

Hence $c_a = 0$ at $a^2 = -p_1$ provided $m_1 > 0$, but can never tend to zero as $a^2 \to \pm\infty$ when $s > 0$. In other words, of all the possible forms of $y_0 e^{-\frac{m_1}{p_1+u^2}}$, the only equation which can represent a real frequency distribution is

$$y = y_0 e^{-\frac{m}{p-u^2}},$$

where we must take the range for u from $-\sqrt{p}$ to $+\sqrt{p}$. This solution corresponds to $m_1 < 0$, $\frac{b_0}{b_4} > 0$ and $\frac{b_2}{b_4} < 0$ in the differential equation (5).

In the present paper I shall confine myself only to those equations which satisfy the conditon $c_a = 0$.

Consider the equation

$$y = y_0 \left(\frac{q - u^2}{p + u^2}\right)^{k'},$$

where

$$p = |p_1|,$$
$$q = |q_1|,$$
$$m = |m_1|,$$
$$k' = |k_1'| = \frac{m}{q+p}.$$

I shall take the limits of the range, viz. a, to be defined by the relation $c_a = 0$, and hence in the same way as above we find $a = \pm\sqrt{q}$. The corresponding differential equation is

$$\frac{1}{y}\frac{dy}{du} = \frac{2mu}{pq - (q-p)u^2 + u^4}.$$

Compare this equation with the Equation (5). Hence

$m_1 = m$, and therefore $m_1 > 0$,

$\frac{b_0}{b_4} = pq$, " " $\frac{b_0}{b_4} > 0$,

$\frac{b_2}{b_4} = -(q-p)$, " " $\frac{b_2}{b_4} < 0$,

which gives us the criterion for this type of curve, which I shall call the Type (i) curve. In this way we find the following types of curves and their criterion:

TABLE I.

Type	Form of Curve	Range for u	Criterion		
			b-ratios	m_1	k or k'
(i)	$y=y_0\left(\frac{q-u^2}{p+u^2}\right)^{k'}$	$-\sqrt{q}$ to $+\sqrt{q}$	$\frac{b_2}{b_4}<0$ $\frac{b_0}{b_4}<0$	$m_1>0$	...
(ii)	$y=y_0\left(\frac{p-u^2}{q+u^2}\right)^{k'}$	$-\sqrt{p}$ to $+\sqrt{p}$	$\frac{b_2}{b_4}>0$ $\frac{b_0}{b_4}<0$	$m_1>0$	...
*(iii)	$y=y_0\left(\frac{q-u^2}{p-u^2}\right)^{k}$	$-\sqrt{p}$ to $+\sqrt{p}$	$\frac{b_2}{b_4}<0$ $\frac{b_0}{b_4}>0$	$m_1>0$	$k<1$
*(iv)	$y=y_0\left(\frac{p+u^2}{q-u^2}\right)^{k'}$	$-\sqrt{q}$ to $+\sqrt{q}$	$\frac{b_2}{b_4}<0$ $\frac{b_0}{b_4}<0$	$m_1<0$	$k'<1$
*(v)	$y=y_0\left(\frac{q+u^2}{p-u^2}\right)^{k'}$	$-\sqrt{p}$ to $+\sqrt{p}$	$\frac{b_2}{b_4}>0$ $\frac{b_0}{b_4}<0$	$m_1<0$	$k'<1$
(vi)	$y=y_0\left(\frac{p-u^2}{q-u^2}\right)^{k}$	$-\sqrt{p}$ to $+\sqrt{p}$	$\frac{b_2}{b_4}<0$ $\frac{b_0}{b_4}>0$	$m_1<0$	...
(vii)	$y=y_0 e^{-\frac{m}{p-u^2}}$ †	$-\sqrt{p}$ to $+\sqrt{p}$	$\frac{b_2}{b_4}<0$ $\frac{b_0}{b_4}>0$	$m_1<0$	$\left(\frac{b_2}{b_4}\right)^2=4\frac{b_0}{b_4}$

* The Types (iii), (iv) and (v) are u-shaped curves. † $m=|m_1|$.

Note that for the Types (i)—(vi) we must also have the condition $\left(\frac{b_2}{b_4}\right)^2>4\frac{b_0}{b_4}$, so that p and q may be real.

The general differential equation for the Types (i)—(vi) is

$$\frac{1}{y}\frac{dy}{du}=\frac{2m_1u}{p_1q_1+(p_1+q_1)u^2+u^4}.$$

Compare this with the Equation (5).

Therefore

$$\left.\begin{aligned} m_1 &= \frac{1}{2b_4} \\ p_1 q_1 &= \frac{b_0}{b_4} \\ p_1 + q_1 &= \frac{b_2}{b_4} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6).$$

Let us express the quantities b_0, b_2 and b_4, and therefore m_1, $p_1 q_1$ and $p_1 + q_1$, in terms of β_2, β_4 and β_6.

Put $c_a = 0$ in (3) and the difference equation reduces to

$$(2s+1)\,\mu_{2s} b_0 \mu_2 + (2s+3)\, b_2 \mu_{2s+2} + (2s+5)\,\frac{b_4}{\mu_2}\,\mu_{2s+4} = -\,\mu_{2s+2}.$$

Put $s=2$: $\quad b_0\mu_2 + 3b_2\mu_2 + 5\,\dfrac{b_4}{\sigma^2}\,\mu_4 = -\,\mu_2,$

$s=1$: $\quad 3b_0\mu_2^2 + 5b_2\mu_4 + 7\,\dfrac{b_4}{\mu_2}\,\mu_6 = -\,\mu_4,$

$s=2$: $\quad 5b_0\mu_2\mu_4 + 7b_2\mu_6 + 9\,\dfrac{b_4}{\mu_2}\,\mu_8 = -\,\mu_6.$

Divide the equations above by μ_2, μ_2^2, μ_2^3 respectively. Write

$$\beta_2 = \frac{\mu_4}{\mu_2^2}, \quad \beta_4 = \frac{\mu_6}{\mu_2^3} \text{ and } \beta_6 = \frac{\mu_8}{\mu_2^4}.$$

We obtain

$$\left.\begin{aligned} b_0 + 3b_2 + 5b_4\beta_2 &= -1 \\ 3b_0 + 5b_2\beta_2 + 7b_4\beta_4 &= -\beta_2 \\ 5b_0\beta_2 + 7b_2\beta_4 + 9b_4\beta_6 &= -\beta_4 \end{aligned}\right\}.$$

Solving, for b_0, b_2 and b_4, we get

$$\left.\begin{aligned} b_0 &= \frac{2\,\{14\beta_4^2 - 9\beta_2\beta_6 - 5\beta_2^2\beta_4\}}{9\beta_6(5\beta_2 - 9) - 49\beta_4^2 + 210\beta_2\beta_4 - 125\beta_2^3} \\ b_2 &= \frac{\{9\beta_6(3-\beta_2) + 7\beta_4^2 - 50\beta_2\beta_4 + 25\beta_2^3\}}{\{9\beta_6(5\beta_2 - 9) - 49\beta_4^2 + 210\beta_2\beta_4 - 125\beta_2^3\}} \\ b_4 &= \frac{2\,\{5\beta_2^2 - \beta_4(6-\beta_2)\}}{9\beta_6(5\beta_2 - 9) - 49\beta_4^2 + 210\beta_2\beta_4 - 125\beta_2^3} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots (7).$$

The solutions are valid, provided that

$$\Delta = 9\beta_6(5\beta_2 - 9) - 49\beta_4^2 + 210\beta_2\beta_4 - 125\beta_2^3$$

is not zero.

If $\Delta = 0$, we solve for $\quad \dfrac{b_0}{b_4}, \quad \dfrac{b_2}{b_4}, \quad \dfrac{1}{b_4},$

and find

$$\left.\begin{aligned} \frac{b_0}{b_4} &= \frac{14\beta_4^2 - 9\beta_2\beta_6 - 5\beta_2^2\beta_4}{5\beta_2^2 - (6-\beta_2)\beta_4} \\ \frac{b_2}{b_4} &= \frac{9\beta_6(3-\beta_2) + 7\beta_4^2 - 50\beta_2\beta_4 + 25\beta_2^3}{2\,\{5\beta_2^2 - (6-\beta_2)\beta_4\}} \\ \frac{1}{b_4} &= \frac{9\beta_6(5\beta_2 - 9) - 49\beta_4^2 + 210\beta_2\beta_4 - 125\beta_2^3}{2\,\{5\beta_2^2 - (6-\beta_2)\beta_4\}} \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots (8).$$

Since $\Delta = 0$, i.e.

$$9\beta_6(5\beta_2 - 9) - 49\beta_4^2 + 210\beta_2\beta_4 - 125\beta_2^3 = 0,$$

$\frac{1}{b_4} = 0$, but $\frac{b_0}{b_4}$ and $\frac{b_2}{b_4}$ remain finite.

Hence our differential equation reduces to

$$\frac{1}{y}\frac{dy}{dx} = 0.$$

The integral is now $y =$ constant, or the frequency curve is a rectangle, which I shall call the Type (viii) curve.

The solutions (8) above are valid only if $5\beta_2^2 - (6 - \beta_2)\beta_4 \neq 0$. If

$$5\beta_2^2 - (6 - \beta_2)\beta_4 = 0 \quad \text{and} \quad 9\beta_6(5\beta_2 - 9) - 49\beta_4^2 + 210\beta_2\beta_4 - 125\beta_2^3 = 0$$

are simultaneously true, i.e. if $\beta_4 = \frac{5\beta_2^2}{6 - \beta_2}$ and $\beta_6 = \frac{25\beta_2^3(\beta_2 + 8)}{9(6 - \beta_2)^2}$, each of the expressions

$$5\beta_2^2 - (6 - \beta_2)\beta_4,$$

$$9\beta_6(5\beta_2 - 9) - 49\beta_4^2 + 210\beta_2\beta_4 - 125\beta_2^3,$$

$$14\beta_4^2 - 9\beta_2\beta_6 - 5\beta_2^2\beta_4,$$

$$9\beta_6(3 - \beta_2) + 7\beta_4^2 - 50\beta_2\beta_4 + 25\beta_2^3$$

becomes zero and we cannot find the values of b_1, b_2 and b_4, or the corresponding type of frequency distribution from the difference equation. In Section *h* (β) I shall show that the frequency type corresponding to the values of

$$\left.\begin{aligned} \beta_4 &= \frac{5\beta_2^2}{6 - \beta_2} \\ \beta_6 &= \frac{25\beta_2^3(\beta_2 + 8)}{9(6 - \beta_2)^2} \end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

is a special form of the Type (vii) curve, viz. $y = y_0 e^{p - x^2/\sigma^2}$.

(*b*) *General Discussion of the Constants of the Curve.*

The following values were found:

$$p_1 q_1 = \frac{b_0}{b_4} = -\frac{2(9\beta_2\beta_6 - f_0)}{2K},$$

where $f_0 = 14\beta_4^2 - 5\beta_2^2\beta_4$ and $K = 5\beta_2^2 - (6 - \beta_2)\beta_4$,

$$(q_1 + p_1) = \frac{b_2}{b_4} = \frac{9\beta_6(3 - \beta_2) - f_2}{2K},$$

where $f_2 = -(7\beta_4^2 - 50\beta_2\beta_4 + 25\beta_2^3)$,

$$m_1 = \frac{1}{2b_4} = \frac{9\beta_6(5\beta_2 - 9) - f_4}{4K},$$

where $f_4 = 49\beta_4^2 - 210\beta_2\beta_4 + 125\beta_2^3$.

I have divided the curves into the different types according to

(i) the sign of $\frac{b_0}{b_4}$, $\frac{b_2}{b_4}$, and m_1, hence the sign of p_1q_1, p_1+q_1, and m_1; in other words we want the signs of the expressions

$$\left.\begin{aligned} K &= 5\beta_2^2-(6-\beta_2)\beta_4 \\ F_0 &= 9\beta_2\beta_6-f_0 \\ F_2 &= 9\beta_6(3-\beta_2)-f_2 \\ F_4 &= 9\beta_6(5\beta_2-9)-f_4 \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(10);$$

(ii) whether p_1 and q_1 are equal, real or imaginary; hence we require the sign of

$$(q_1-p_1)^2=\frac{81\beta_6^2(3-\beta_2)^2-18\beta_6 b+c}{4K^2},$$

where $b=(3-\beta_2)f_2-8\beta_2K$, $c=f_2^2-16f_0K$, and $L=7\beta_4(\beta_2-3)+\beta_2^2(19-5\beta_2)$. On simplification

$$(q_1-p_1)^2=\frac{81(\beta_2-3)^2}{4K^2}\left\{\beta_6-\frac{b+4|K|\sqrt{L}}{9(\beta_2-3)^2}\right\}\left\{\beta_6-\frac{b-4|K|\sqrt{L}}{9(\beta_2-3)^2}\right\}.$$

I shall write this equation as

$$F_6\equiv\{F_6'\}\{F_6''\}\frac{81(\beta_2-3)^2}{4K^2},$$

where $$F_6'=\beta_6-\frac{b+4|K|\sqrt{L}}{9(3-\beta_2)^2},\quad F_6''=\beta_6-\frac{b-4|K|\sqrt{L}}{9(3-\beta_2)^2};$$

(iii) whether k or $k'\gtreqless 1$, hence we require the sign of $[m_1-(q_1-p_1)]$. The zone where $c_a\neq 0$, where p_1q_1, p_1+q_1, and m_1 have other values than those we found, is bounded by $F_6=0$ and $m_1=(q_1-p_1)$. We therefore want to write the condition $[m_1-(q_1-p_1)]\gtreqless 0$ in a workable form. To avoid the difficulty of the square root and also of sign, I shall consider $F_8=m_1^2-(q_1-p_1)^2$.

On simplification we find:

$$F_8=\frac{243\beta_6^2(\beta_2-1)(7\beta_2-15)-18\beta_6B+C}{16K^2}$$

$$=\frac{243(\beta_2-1)(7\beta_2-15)}{16K^2}\{F_8'\}\{F_8''\},$$

where

$$F_8'=\beta_6-\frac{B+4|K|\sqrt{Q}}{27(\beta_2-1)(7\beta_2-15)},$$

$$F_8''=\beta_6-\frac{B-4|K|\sqrt{Q}}{27(\beta_2-1)(7\beta_2-15)},$$

$$B=(5\beta_2-9)f_4-4b,$$

$$C=f_4^2-4c,$$

$$Q=49\beta_4^2-14\beta_4(37\beta_2-30)+7\beta_2^2(60\beta_2-53).$$

Hence $F_x=0$ ($x=0, 2, 4$) and $F_y=0$ ($y=6, 8$) represent surfaces of revolution in the β_2, β_4, β_6 space such that sections by the plane $\beta_2=$ constant of $F_x=0$ represent parabolas and of $F_y=0$ two intersecting parabolas.

The position of any point $(\beta_2, \beta_4, \beta_6)$ with respect to these surfaces of revolution determines the signs of the expressions F_x and F_y, and the signs of p_1q_1, p_1+q_1, m_1, $(q_1-p_1)^2$, and $[m_1^2-(q_1-p_1)^2]$. Hence, theoretically, a model of these surfaces of revolution will enable us to determine what type of curve corresponds to a given set of values of β_2, β_4 and β_6.

(*c*) *The Transition Types.*

From the previous Section it is clear that when β_2, β_4 and β_6 satisfy $F_x=0$ $(x=0, 2, 4, 6, 8)$ either b_0, b_2 or b_4 vanishes or there is a relation between these b's, and the general differential equation is simplified. To each variation in the form of the differential equation there corresponds an integral. Here again some of the types so obtained are not admissible because we are confining ourselves to those types of frequency curves which satisfy $c_a=0$. The discussion is similar to that given in Section (*a*) and is omitted here for the sake of brevity. The results are summarised below in Table II.

Notes. (1) When $\beta_4=\dfrac{5\beta_2^2}{6-\beta_2}$, $b_4=0$ and the general differential equation reduces to

$$\frac{1}{y}\frac{dy}{du}=\frac{u}{b_0+b_2u^2}.$$

This is the equation Professor Pearson integrates to obtain his symmetrical types. Let us see if by proceeding to the limit as $\beta_4\to\dfrac{5\beta_2^2}{6-\beta_2}$ we get the same values as he deduces for his curve.

$$b_2=\frac{9\beta_6(3-\beta_2)-f_2}{9\beta_6(5\beta_2-9)-f_4}=\frac{9\beta_6(3-\beta_2)+7\beta_4^2-50\beta_2\beta_4+25\beta_2^3}{9\beta_6(5\beta_2-9)-49\beta_4^2+210\beta_2\beta_4-125\beta_2^3},$$

therefore

$$\lim_{\beta_4\to 7\frac{5\beta_2^2}{6-\beta_2}} b_2=\frac{\dfrac{3-\beta_2}{(6-\beta_2)^2}\{9\beta_6(6-\beta_2)^2-25\beta_2^3(\beta_2+8)\}}{\dfrac{5\beta_2-9}{(6-\beta_2)^2}\{9\beta_6(6-\beta_2)^2-25\beta_2^3(\beta_2+8)\}}$$

$$=\frac{3-\beta_2}{5\beta_2-9}, \qquad \left(\beta_6\neq\frac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2}\right).$$

Similarly

$$\lim_{\beta_4\to 7\frac{5\beta_2^2}{6-\beta_2}} b_0=-\frac{2\beta_2}{5\beta_2-9}.$$

It would seem as if we should have Professor Pearson's types at every point along the curve $\beta_4=\dfrac{5\beta_2^2}{6-\beta_2}$, no matter what the value of β_6 is. But this is not so. When $\beta_4=\dfrac{5\beta_2^2}{6-\beta_2}$, $b_4=0$, b_2 and b_0 can be expressed in terms of β_2 and are fixed for constant β_2. We get a relation between β_6, β_4 and β_2, and find

$$\beta_6=\frac{35\beta_2^3}{(6-\beta_2)(9-2\beta_2)}.$$

TABLE II. *Table showing the various Types of Frequency Curves.*

	Type	Equation of Curve	Limits of the Range	Differential Equation	Criterion m_1	Criterion p_1q_1	Criterion q_1+p_1	Criterion $(q_1-p_1)^2$	Criterion $m_1^2-(q_1-p_1)^2$	Remarks
Main Types	(i)	$y=y_0\left[\frac{q-u^2}{p+u^2}\right]^{k'}$	$\pm\sqrt{q}$	$\frac{1}{y}\frac{dy}{du}=\frac{2mu}{-pq-(q-p)u^2+u^4}$	+	−	−	+	+ or −	
	(ii)	$y=y_0\left[\frac{p-u^2}{q+u^2}\right]^{k'}$	$\pm\sqrt{p}$	$\frac{1}{y}\frac{dy}{du}=\frac{2mu}{-pq+(q-p)u^2+u^4}$	+	−	+	+	+ or −	
	(iii)	$y=y_0\left[\frac{q-u^2}{p-u^2}\right]^{k}$	$\pm\sqrt{p}$	$\frac{1}{y}\frac{dy}{du}=\frac{2mu}{pq-(q+p)u^2+u^4}$	+	+	−	+	−	
	(iv)	$y=y_0\left[\frac{p+u^2}{q-u^2}\right]^{k'}$	$\pm\sqrt{q}$	$\frac{1}{y}\frac{dy}{du}=\frac{-2mu}{-pq-(q-p)u^2+u^4}$	−	−	−	+	−	
	(v)	$y=y_0\left[\frac{q+u^2}{p-u^2}\right]^{k'}$	$\pm\sqrt{p}$	$\frac{1}{y}\frac{dy}{du}=\frac{-2mu}{-pq+(q-p)u^2+u^4}$	−	−	+	+	−	
	(vi)	$y=y_0\left[\frac{p-u^2}{q-u^2}\right]^{k}$	$\pm\sqrt{p}$	$\frac{1}{y}\frac{dy}{du}=\frac{-2mu}{pq-(q+p)u^2+u^4}$	−	+	−	+	+ or −	
Transition Types	(vii)	$y=y_0\,e^{-\frac{m}{p-u^2}}$	$\pm\sqrt{p}$	$\frac{1}{y}\frac{dy}{du}=\frac{-2mu}{p^2-2pu^2+u^4}$	−	+	−	0	+ or −	β_2, β_4 and β_6 satisfy $F_6=0$
	(viii)	$y=y_0=$const.	...	$\frac{1}{y}\frac{dy}{du}=0$	For all valves of p_1 and q_1 provided p_1q_1, p_1+q_1 and m_1 are not simultaneously zero. β_2, β_4 and β_6 satisfy $F_4=0$ or simultaneously $F_4=0$ and $F_2=0$, $F_4=0$ or $F_8=0$, $F_6=0$ and $F_8=0$.					
	(ix)	$y=y_0\left[\frac{p-u^2}{p+u^2}\right]^{\frac{m}{2p}}$	$\pm\sqrt{p}$	$\frac{1}{y}\frac{dy}{du}=\frac{2mu}{-p^2+u^4}$	+	−	0	+	+ or −	β_2, β_4 and β_6 satisfy $F_2=0$
	(x)	$y=y_0\left[\frac{p+u^2}{p-u^2}\right]^{\frac{m}{2p}}$	$\pm\sqrt{p}$	$\frac{1}{y}\frac{dy}{du}=\frac{-2mn}{-p^2+u^4}$	−	−	0	+	−	
	(xi)	$y=y_0\left[\frac{q-u^2}{u^2}\right]^{\frac{m}{q}}$	$\pm\sqrt{q}$	$\frac{1}{y}\frac{dy}{du}=\frac{+2mu}{-qu^2+u^4}$	+	0	−	+	$m<q$	β_2, β_4 and β_6 satisfy $F_0=0$
	(xii)	$y=y_0\left[\frac{u^2}{q-u^2}\right]^{\frac{m}{q}}$	$\pm\sqrt{q}$	$\frac{1}{y}\frac{dy}{du}=\frac{-2mu}{-qu^2+u^4}$	−	0	−	+	$m<q$	
	(xiii)	$y=y_0\left[\frac{q-u^2}{p+u^2}\right]$	$\pm\sqrt{q}$	$\frac{1}{y}\frac{dy}{du}=\frac{2(p+q)u}{-pq-u^2(q-p)+u^4}$	+	−	−	+	0	β_2, β_4 and β_6 satisfy $F_8=0$
	(xiv)	$y=y_0\left[\frac{p-u^2}{q+u^2}\right]$	$\pm\sqrt{p}$	$\frac{1}{y}\frac{dy}{du}=\frac{2(p+q)u}{-qp+(q-p)u^2+u^4}$	+	−	+	+	0	
	(xv)	$y=y_0\left[\frac{p-u^2}{q-u^2}\right]$	$\pm\sqrt{p}$	$\frac{1}{y}\frac{dy}{du}=\frac{-2(q-p)u}{pq-(p+q)u^2+u^4}$	−	+	−	+	0	
	(xvi)	$y=y_0\,(p^2-u^2)^m$	$\pm p$	$\frac{1}{y}\frac{dy}{du}=\frac{-2mu}{p^2-u^2}$	$\beta_1=\frac{5\beta_2^2}{6-\beta_2}$					Type II-curve
	(xvii)	$y=y_0\,(p^2+u^2)^{-m}$	$\pm\infty$	$\frac{1}{y}\frac{dy}{du}=\frac{-2mu}{p^2+u^2}$						Type VII-curve
	(xviii)	$y=y_0\,(p^2-u^2)^{-m}$	$\pm p$	$\frac{1}{y}\frac{dy}{du}=\frac{2mu}{p^2-u^2}$	$\beta_6=\frac{35\beta_2^3}{(9-2\beta_2)(6-\beta_2)}$					Type II (u-shaped)
	(xix)	$y=y_0\,e^{-\frac{1}{2}u^2}$	$\pm\infty$	$\frac{1}{y}\frac{dy}{du}=-u$	$\beta_2=3,\ \beta_4=15,\ \beta_6=105$					Normal curve
	(xx)	$y=y_0\left[\frac{p-u^2}{p+u^2}\right]$	$\pm\sqrt{p}$	$\frac{1}{y}\frac{dy}{du}=\frac{4pu}{p^2-u^2}$	−	−	0	+	0	β_2, β_4 and β_6 simultaneously satisfy $F_2=0$ and $F_8=0$

$m=|m_1|$, $p=|p_1|$, $q=|q_1|$, $k=\frac{m}{q-p}$, $k'=\frac{m}{q+p}$.

Hence there is only one point on the curve $\beta_4 = \frac{5\beta_2^2}{6-\beta_2}$ for each β_2 which corresponds to Professor Pearson's symmetrical curves. At other points we can have other types only.

The normal curve is a special case of the 2nd order symmetrical curves and corresponds to the point $\beta_2 = 3$, $\beta_4 = 15$ and $\beta_6 = 105$.

(2) Consider the equation $y = y_0 e^{\frac{m_1}{u^2}}$.

This equation is obtained by integrating the differential equation

$$\frac{1}{y}\frac{dy}{du} = \frac{2m_1 u}{u^4} = \frac{2m_1}{u^3}.$$

Hence it is the type we find along the curves of section of the surfaces

(i) $F_0 = 0$ and $F_2 = 0$, (ii) $F_0 = 0$ and $F_6 = 0$, (iii) $F_2 = 0$ and $F_6 = 0$.

Note that since $F_0 = p_1 q_1$, $F_2 = q_1 + p_1$, $F_6 = (q_1 - p_1)^2$, the surfaces $F_0 = 0$, $F_2 = 0$ and $F_6 = 0$ in general cut each other where $p_1 = 0 = q_1$, and hence we have a common curve of intersection.

For the equation above

$$c_a = \text{const.} \times a^{2s+4} e^{\pm \frac{m\sigma^2}{a^2}},$$

which is never zero and accordingly the equation

$$y = y_0 e^{\pm \frac{m\sigma^2}{x^2}}$$

does not represent a complete frequency distribution.

(3) We can write the Type (xiii) curve as

$$y = y_0 \left[\frac{(q+p)\sigma^2}{p\sigma^2 + x^2} - 1\right], \text{ i.e. } y + y_0 = y_0 (q+p)\sigma^2 \frac{1}{p\sigma^2 + x^2}.$$

Hence the Type (xiii) curve is the part above the line $y = y_0$ of Pearson's Type VII curve $y = y_0 \frac{(q+p)\sigma^2}{p\sigma^2 + x^2}$. Similarly the Types (xiv) and (xv) represent the parts above the line $y = y_0$ of the curve $y = y_0 \frac{(p+q)\sigma^2}{q\sigma^2 + x^2}$ and the inverted u-shaped curve

$$y = y_0 \frac{(q-p)\sigma^2}{q\sigma^2 - x^2}.$$

(d) and (e). The Inequalities satisfied by β_2, β_4 and β_6; Note on the Surfaces $F_6 = 0$ and $F_8 = 0$.

Before we can discuss the intersections of the transition surfaces with each other we must know something about the relations between β_2, β_4 and β_6, and also when the branches of $F_6 = 0$ and $F_8 = 0$ become imaginary.

In this way we introduce the further conditions, viz.

(i) $\beta_6 > \beta_2\beta_4 > \beta_2^3$ and $\beta_4 > \beta_2^2$*,

and (ii) $L > 0$, $Q > 0$.

It is easy to show that

I. (i) for $\beta_2 > 3$, $F_6' = 0$ and $F_6'' = 0$ cut at $\beta_4 = \dfrac{5\beta_2^2}{6-\beta_2}$ and are always real;

(ii) for $\beta_2 < 3$, $F_6' = 0$ and $F_6'' = 0$ cut at $\beta_4 = \dfrac{5\beta_2^2}{6-\beta_2}$ and again at

$$\beta_4 = \frac{\beta_2^2(19-5\beta_2)}{7(3-\beta_2)}.$$

For values of $\beta_4 > \dfrac{\beta_2^2(19-5\beta_2)}{7(3-\beta_2)}$ the branches $F_6' = 0$ and $F_6'' = 0$ become imaginary.

II. (i) for $\beta_2 > \frac{15}{7}$, $F_8' = 0$ and $F_8'' = 0$ cut at $\beta_4 = \dfrac{5\beta_2^2}{6-\beta_2}$ and are always real;

(ii) for $\beta_2 < \frac{15}{7}$, $F_8' = 0$ and $F_8'' = 0$ cut at $\beta_4 = \dfrac{5\beta_2^2}{6-\beta_2}$, $\beta_4 = d'$, $\beta_4 = d''$, the curves becoming imaginary for values of β_4 between d' and d'', where

$$d' = \tfrac{1}{7}[(37\beta_2 - 30) + 2(\beta_2 - 1)\sqrt{15(15-7\beta_2)}],$$
$$d'' = \tfrac{1}{7}[(37\beta_2 - 30) - 2(\beta_2 - 1)\sqrt{15(15-7\beta_2)}].$$

(*f*) *The Intersections of the Transition Surfaces with each other.*

Let ϕ_x denote the value of β_6 on the surface $F_x = 0$. Then by finding the values of β_4 in terms of β_2, which make $(\phi_r - \phi_s) = 0$, $\left(\begin{matrix} r \\ s \end{matrix} = 0, 2, 4, 6, 8\right)$, and by substitution in $F_r = 0$ or $F_s = 0$ the value of β_6, also expressed in terms of β_2, we can determine the curve of section of the surfaces $F_r = 0$ and $F_s = 0$.

We have found

$$\phi_0 = \frac{f_0}{9\beta_2} = (14\beta_4^2 - 5\beta_2^2\beta_4)/9\beta_2,$$

$$\phi_2 = \frac{f_2}{9(3-\beta_2)} = -\{7\beta_4^2 - 50\beta_2\beta_4 + 25\beta_2^3\}/9(3-\beta_2),$$

$$\phi_4 = \frac{f_4}{9(5\beta_2 - 9)} = \{49\beta_4^2 - 210\beta_2\beta_4 + 125\beta_2^3\}/9(5\beta_2 - 9),$$

$$\phi_6'' = \frac{b - 4|K|\sqrt{L}}{9(3-\beta_2)^2},$$

$$\phi_8'' = \frac{B - 4|K|\sqrt{Q}}{27(\beta_2 - 1)(7\beta_2 - 15)},$$

* J. Shohat, *Biometrika*, Vol. XXI. For a discontinuous distribution the proof is straightforward and similar to that given by the Editorial Note on Shohat's memoir.

where
$$K = 5\beta_2^2 - (6-\beta_2)\beta_4,$$
$$b = (3-\beta_2)f_2 - 8\beta_2 K,$$
$$L = \beta_2^2(19-5\beta_2) - 7(3-\beta_2)\beta_4,$$
$$B = (5\beta_2-9)f_4 - 4b,$$
$$Q = 49\beta_4^2 - 14\beta_4(37\beta_2-30) + 7\beta_2^2(60\beta_2-53).$$

(α) The following table gives the values of β_4 and β_6, in terms of β_2, at the intersection of F_0, F_2 and $F_4=0$ with each other:

TABLE III. *Values of β_4 and β_6.*

Surfaces	$\phi_r - \phi_s$	Values of β_4	Values of β_6
$F_0=0$ and $F_4=0$	$\frac{(21\beta_4-25\beta_2^2)(5\beta_2^2-\overline{6-\beta_2}\beta_4)}{9\beta_2(5\beta_2-9)}$	$5\beta_2^2/6-\beta_2$ $25\beta_2^2/21$	$25\beta_2^3(\beta_2+8)/9(6-\beta_2)^2$ $125\beta_2^3/81$
$F_2=0$ and $F_4=0$	$\frac{2(7\beta_4-15\beta_2)(5\beta_2^2-\overline{6-\beta_2}\beta_4)}{9(3-\beta_2)(5\beta_2-9)}$	$5\beta_2^2/6-\beta_2$ $15\beta_2/7$	$25\beta_2^3(\beta_2+8)/9(6-\beta_2)^2$ $25\beta_2^2/9$
$F_2=0$ and $F_0=0$	$\frac{(7\beta_4-5\beta_2^2)(5\beta_2^2-\overline{6-\beta_2}\beta_4)}{9\beta_2(3-\beta_2)}$	$\frac{5\beta_2^2}{6-\beta_2}$ *	$\frac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2}$

(β) The discussion of the values of β_4 and β_6, in terms of β_2, at the intersections of $F_x=0$ and $F_y=0$ $\left(\begin{matrix}x=0,\,2,\,4\\ y=6,\,8\end{matrix}\right)$ is more complicated. We know that all these surfaces have a common curve of section, viz. $\beta_2=\beta_2$, $\beta_4=\frac{5\beta_2^2}{6-\beta_2}$, $\beta_6=\frac{25\beta_2^3(\beta_2+8)}{9(6-\beta^2)^2}$ and this enables us to simplify the lengthy algebra since $(\phi_r-\phi_s)\left(\begin{matrix}r=0,2,4,6,8\\ s\end{matrix}\right)$ must have $\beta_4=\frac{5\beta_2^2}{6-\beta_2}$ as a factor. Furthermore, the surfaces $F_4=0$, $F_6=0$ and $F_8=0$ have a common curve of section, which by considering the expression for $(\phi_6-\phi_4)$ is found to be

$$\beta_2=\beta_2,\quad \beta_4=\tfrac{1}{7}\{9(5\beta_2-6)\pm 2(9-5\beta_2)^{\frac{3}{2}}\},\quad \beta_6=\tfrac{1}{3}\{\pm 8(9-5\beta_2)^{\frac{3}{2}} + (180\beta_2-25\beta_2^2-216)\}.$$

Similarly the curve of section of $F_8=0$ and $F_0=0$ is

$$\beta_2=\beta_2,\quad \beta_4=\frac{15\beta_2^2}{7},\quad \beta_6=\frac{125\beta_2^3}{21},$$

and finally that of $F_8=0$ and $F_2=0$

$$\beta_2=\beta_2,\quad \beta_4=\tfrac{1}{7}\{(7\beta_2+24)+4|\beta_2-3|\sqrt{5\beta_2+4}\},$$
$$\beta_6=\tfrac{1}{21}\{(85\beta_2^2+16\beta_2-128)+16(3\beta_2-4)(5\beta_2+4)^{\frac{1}{2}}\}.$$

* Note $\beta_4>\beta_2^2$ and therefore $7\beta_4>5\beta_2^2$ for all values of β_2 and β_4.

The numerators of the corresponding expressions $(\phi_r - \phi_s)$ each contain a square root and in obtaining the results above we may have introduced inadmissible values by squaring; furthermore the values of β_4 and β_6 must satisfy the inequalities

(i) $\beta_6 > \beta_2\beta_4 > \beta_2^3$, $\beta_4 > \beta_2^2$,

(ii) $L > 0$ and $Q > 0$.

After some very lengthy algebra, I found for which values of β_2 the above solutions are admissible and interpreted the + and − signs of the square roots above. This discussion is given in the original manuscript*.

(*g*) *On the Division of the β_2, β_4, β_6 Space into Zones corresponding to the different Types.*

In the previous Section the value of the expressions $(\phi_r - \phi_s)$ was found and by discussing their sign we are able to find for which values of β_2 and β_4, $\phi_r >$ or $< \phi_s$. This enables us to determine the relative positions of $F_r = 0$ and $F_s = 0$ for any value of β_2, β_4 and β_6 and also to discuss theoretically the curves of section of the transition surfaces corresponding to any given value of β_2. The introduction of the surfaces $F_6 = 0$ and $F_8 = 0$ makes the classification of the different systems of curves of section very complicated and I have confined myself to the surfaces F_0, F_2 and $F_4 = 0$. Their curves of section can be divided into three main classes, according as $\beta_2 < 1{\cdot}8$, $1{\cdot}8 < \beta_2 < 3$, $\beta_2 > 3$, as shown in Diagram I, Figs. I, II, and III. The types of curve in the various zones are found by discussing the signs of F_0, F_2 and F_4 for the corresponding values of β_2 and β_4. These figures were of great use in the construction of Diagrams XII to XX, serving not only as a check, but also enabling us to determine very rapidly the types of curve corresponding to the different zones.

Note. The curve $F_2 = 0$ in Diagram I, Fig. III, has now no meaning because points on it do not satisfy the condition $c_a = 0$. I have reproduced it for the sake of uniformity.

(*h*) *The Constants for the main Types.*

The constants for the main types of frequency curves are expressed in terms of β_2, β_4 and β_6, viz.

$$m_1 = \frac{9}{4K}(5\beta_2 - 9)(\beta_6 - \phi_4),$$

$$q_1 p_1 = -\frac{9\beta_2}{K}(\beta_6 - \phi_0),$$

$$q_1 + p_1 = \frac{9(3-\beta_2)}{2K}(\beta_6 - \phi_2),$$

$$\frac{m_1}{q_1 - p_1} = \pm\sqrt{1 + \frac{m_1^2 - (q_1 - p_1)^2}{(q_1 - p_1)^2}}$$

$$= \pm\sqrt{1 + \frac{3(\beta_2 - 1)(7\beta_2 - 15)}{4(\beta_2 - 3)^2}\frac{(\beta_6 - \phi_8')(\beta_6 - \phi_8'')}{(\beta_6 - \phi_6')(\beta_6 - \phi_6'')}}.$$

* Copy in the Library, University of London.

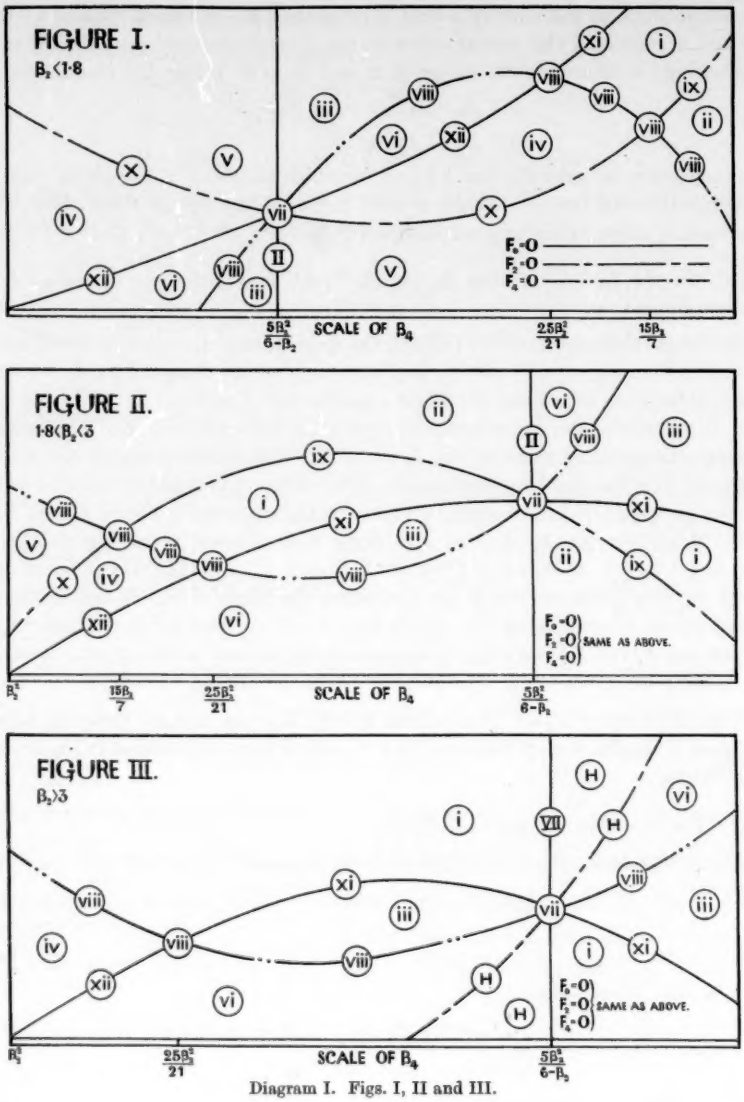


Diagram I. Figs. I, II and III.

The last formula which gives $k=\dfrac{m}{q-p}$ and also $k'=\dfrac{m}{q+p}$ is too complicated to be of any use in practice. It is far simpler to find m_1 and (q_1-p_1) from the formulae (6) and (7) and then $\dfrac{m_1}{q_1-p_1}$ by division.

Let us now find the constants of integration of each curve. Let the total frequency be N, $f(x)$ the frequency curve which is symmetrical, $-b$ and $+b$ the limits of frequency, then

$$N = \int_{-b}^{+b} f(x)\,dx = 2\int_{0}^{+b} f(x)\,dx.$$

By expanding $f(x)$, we easily find the following expressions for y_0:

TABLE IV. *Values of* y_0.

Type	Integral for N	Expression for y_0
(i)	$N = 2y_0\int_0^{\sqrt{q}.\sigma}\left(\frac{1-\frac{x^2}{q\sigma^2}}{1+\frac{x^2}{p\sigma^2}}\right)^{k'} dx$	$y_0 = \frac{N}{\sigma\sqrt{p}}\Big/\left\{\sum_{r=0}^{\infty}(-1)^r\left(\frac{p}{q}\right)^r \frac{k'!}{(k'-r)!\,r!}\,I_r\right\}$ *
(ii)	$N = 2y_0\int_0^{\sigma\sqrt{p}}\left(\frac{1-\frac{x^2}{p\sigma^2}}{1+\frac{x^2}{q\sigma^2}}\right)^{k'} dx$	$y_0 = \frac{N}{\sigma\sqrt{p}}\,\frac{\Gamma(\frac{3}{2}+k')}{\sqrt{\pi}\,\Gamma(1+k')\,F(k', \frac{1}{2}, k'+\frac{3}{2}, -p/q)}$
(iii)	$N = 2y_0\int_0^{\sigma\sqrt{p}}\left(\frac{1-\frac{x^2}{q\sigma^2}}{1-\frac{x^2}{p\sigma^2}}\right)^{k} dx$	$y_0 = \frac{N}{\sigma\sqrt{p}}\,\frac{\Gamma(\frac{3}{2}-k)}{\sqrt{\pi}\,\Gamma(1-k)\,F(-k, \frac{1}{2}, \frac{3}{2}-k, p/q)}$
(iv)	$N = 2y_0\int_0^{\sigma\sqrt{q}}\left(\frac{1+\frac{x^2}{p\sigma^2}}{1-\frac{x^2}{q\sigma^2}}\right)^{k'} dx$	$y_0 = \frac{N\sqrt{q}\left(\frac{p}{q}\right)^{k'}}{\sigma(p+q)}\,\frac{1}{I'}$ †
(v)	$N = 2y_0\int_0^{\sigma\sqrt{p}}\left(\frac{1+\frac{x^2}{q\sigma^2}}{1-\frac{x^2}{p\sigma^2}}\right)^{k'} dx$	$y_0 = \frac{N}{\sigma\sqrt{p}}\,\frac{\Gamma(\frac{3}{2}-k')}{\sqrt{\pi}\,\Gamma(1-k')\,F(-k', \frac{1}{2}, \frac{3}{2}-k', -p/q)}$
(vi)	$N = 2y_0\int_0^{\sigma\sqrt{p}}\left(\frac{1-\frac{x^2}{p\sigma^2}}{1-\frac{x^2}{q\sigma^2}}\right)^{k} dx$	$y_0 = \frac{N}{\sigma\sqrt{p}}\,\frac{\Gamma(\frac{3}{2}+k)}{\sqrt{\pi}\,\Gamma(1+k)\,F(k, \frac{1}{2}, \frac{3}{2}+k, p/q)}$

* Reduction formula is

$$I_r = -\frac{l^{r-\frac{1}{2}}(1-l)^{k'-r-\frac{1}{2}}}{k'-r-\frac{1}{2}} + \frac{r-\frac{1}{2}}{k'-r-\frac{1}{2}}\,I_{r-1}, \quad I_0 = \beta_l(\tfrac{1}{2}, k'-\tfrac{1}{2}) \text{ and } l = \frac{q}{p+q}.$$

† $$I' = I_0' + \frac{1}{2}\frac{p}{q}\,I_1' + \frac{1}{2}\cdot\frac{3}{2}\cdot\frac{1}{2!}\frac{p^2}{q^2}I_2' + \ldots,$$

where

$I_0' = B_{q/q+p}(1-k', k'+\frac{1}{2})$, an incomplete B-function, $I_r' = -\dfrac{l^{r-k'}(1-l)^{k'+\frac{1}{2}-r}}{k'+\frac{1}{2}-r} + \dfrac{r-k'}{k'+\frac{1}{2}-r}\,I_{r-1}'$.

(For continuation, see Table VI, p. 150.)

The series and the hypergeometrical functions converge rapidly when p/q is small. In other cases it is quicker to use quadrature formula.

The Transition Curves.

There now exists a relation between β_2, β_4 and β_6; by substituting for β_6 in terms of β_2 and β_4 we are able to express the constants of the curve in terms of β_2 and β_4 and thus explain the ambiguity of sign which we find when working from the difference equations. This was necessary

(1) to discuss whether a given point on any of the transition surfaces corresponds to a fourth order Pearson curve consistent with our hypothesis;

(2) to determine the type of curve at all points on the different branches of $F_8 = 0$.

The discussion is long and summarised to some extent in Table V on p. 147. I have excluded from the table the Type (viii) or rectangular distribution, the second order Pearson Curves and the Normal Curve which for convenience I denoted by Types (xvi), (xvii), (xviii) and (xix).

It is interesting to find the values of the constants, expressed in terms of β_2 and β_4, directly from the corresponding difference equation which can be found from first principles or more simply from the general equation (1) on p. 130 by making the appropriate change.

Let us now find the type of curve corresponding to the point I whose ordinates are $\beta_4 = \dfrac{5\beta_2^2}{6-\beta_2}$, $\beta_6 = \dfrac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2}$. Let us approach the point I along either

$$F_6' = 0 \quad \text{or} \quad F_6'' = 0,$$

i.e.
$$p = \frac{-8\beta_2 . K \pm 4|K| . \sqrt{L}}{4(\beta_2-3) . K}$$
$$= \frac{-8\beta_2 . K + 4\sqrt{K^2 . L}}{4K(\beta_2-3)}$$
$$= \frac{-2\beta_2 + \sqrt{L(\beta_4)}}{4(\beta_2-3)},$$

and hence putting $\beta_4 = \dfrac{5\beta_2^2}{6-\beta^2}$ and simplifying,

$$\left.\begin{aligned} P &= \frac{-2\beta_2 + \beta_2\sqrt{\dfrac{(\beta_2-1)(5\beta_2-9)}{6-\beta_2}}}{\beta_2-3}. \\ \text{Similarly the value of } m \text{ when } \beta_4 &= \frac{5\beta_2^2}{6-\beta_2} \text{ is} \\ M &= \frac{(5\beta_2-9)\beta_2\sqrt{\beta_2-1}}{(6-\beta_2)(\beta_2-3)^2}\left[3\sqrt{\beta_2-1} - \sqrt{(5\beta_2-9)(6-\beta_2)}\right], \end{aligned}\right\} \quad \text{.....(11)}$$

on simplifying.

TABLE V. *Values of Constants in terms of β_2 and β_4.*

Type of Curve	Values of the Constants in terms of β_2 and β_4	Remarks
(vii)	$p=\frac{1}{3-\beta_2}(\mp\sqrt{L}+2\beta_2)$ $m=\frac{1}{2(3-\beta_2)^2}\{\mp\sqrt{L}(5\beta_2-9)-[7\beta_4(3-\beta_2)-\beta_2(5\beta_2+9)]\}$	The form of the equation is $y=y_0e^{-\frac{m}{p-u^2}}$. β_2, β_4 and β_6 satisfy $F_6=0$, i.e. $\beta_6=\frac{b\pm4\lvert K\rvert\sqrt{L}}{9(3-\beta_2)^2}$. We must take the + sign if (i) $F_6'=0$ and $K<0$ or (ii) $F_6''=0$ and $K>0$, the − sign if (i) $F_6'=0$ and $K>0$ or (ii) $F_6''=0$ and $K<0$.
(ix)	$p^2=\frac{1}{3-\beta_2}(7\beta_4-5\beta_2)$ $m=\frac{1}{2(3-\beta_2)}(7\beta_4-15\beta_2)$	The form of the equation is $y=y_0\left(\frac{p-u^2}{p+u^2}\right)^{\frac{m}{2p}}$. β_2, β_4 and β_6 satisfy $F_2=0$, i.e. $\beta_6=\frac{1}{9(\beta_2-3)}(7\beta_4^2-50\beta_2\beta_4+25\beta_2^2)$. Since p^2 and m must be positive we must also introduce the conditions $\beta_2<3$ and $\beta_4>\frac{15\beta_2}{7}$.
(x)	$p^2=\frac{1}{3-\beta_2}(7\beta_4-5\beta_2)$ $m=\frac{1}{2(3-\beta_2)}(7\beta_4-15\beta_2)$	The form of the equation is $y=y_0\left(\frac{p+u^2}{p-u^2}\right)^{\frac{m}{2p}}$. β_2, β_4 and β_6 satisfy $F_2=0$. We also have the condition $\beta_2<3$ and $\beta_4<\frac{15\beta_2}{7}$.
(xi)	$q=\frac{1}{2\beta_2}(7\beta_4-5\beta_2^2)$ $m=\frac{1}{4\beta_2}(21\beta_4-25\beta_2^2)$	The form of the curve is $y=y_0\left(\frac{(q-u^2)}{u^2}\right)^{\frac{m}{q}}$. β_2, β_4 and β_6 satisfy $F_0=0$, i.e. $\beta_6=\frac{1}{9\beta_2}(14\beta_2^2-5\beta_2^2\beta_4)$. We must also have $\beta_4>\frac{25\beta_2^2}{21}$ but $<\frac{10\beta_2^2}{7}$*.
(xii)	$q=\frac{7\beta_4-5\beta_2^2}{2\beta_2}$ $m=\frac{1}{4\beta_2}(25\beta_2^2-21\beta_4)$	The form of the curve is $y=y_0\left(\frac{u^2}{q-u^2}\right)^{\frac{m}{q}}$. β_2, β_4 and β_6 satisfy $F_0=0$. The Type (xii) satisfies $c_a=0$ provided $\beta_4<\frac{1}{7}\beta_2^2$.
(xiii)	$p=\frac{1}{2(7\beta_2-15)}\{\pm\sqrt{Q}-(7\beta_4-23\beta_2)\}$ $q=\frac{1}{6(\beta_2-1)}\{\pm\sqrt{Q}+7(\beta_4-\beta_2)\}$	The form of the curve is $y=y_0\left(\frac{q-u^2}{p+u^2}\right)$. β_2, β_4 and β_6 satisfy $F_8=0$, i.e. $\beta_6=\frac{B\pm4\lvert K\rvert\sqrt{Q}}{7(\beta_2-1)(7\beta_2-15)}$. We must take the + sign if (i) $F_8'=0$ and $K>0$ or (ii) $F_8''=0$ and $K<0$, and the − sign if (i) $F_8'=0$ and $K<0$ or (ii) $F_8''=0$ and $K>0$.

* $\frac{m}{q}<\frac{1}{2}$ for finite frequency, which easily reduces to $\beta_4<\frac{10\beta_2^2}{7}$.

TABLE V (*continued*).

Type of Curve	Values of the Constants in terms of β_2 and β_4	Remarks
(xiv)	$p=\frac{1}{6(\beta_2-1)}\{\pm\sqrt{Q}+(7\beta_4-7\beta_2)\}$ $q=\frac{1}{2(7\beta_2-15)}\{\pm\sqrt{Q}-(7\beta_4-23\beta_2)\}$	The form of the curve is $y=y_0\left(\frac{p-u^2}{q+u^2}\right)$. β_2, β_4 and β_6 satisfy $F_8=0$. We must take the + sign if (i) $F_8'=0$ and $K>0$ or (ii) $F_8''=0$ and $K<0$; otherwise the − sign.
(xv)	$p=\frac{1}{2(7\beta_2-15)}\{\pm\sqrt{Q}-(7\beta_4-23\beta_2)\}$ $q=\frac{1}{6(\beta_2-1)}\{\pm\sqrt{Q}+(7\beta_4-7\beta_2)\}$	The form of the curve is $y=y_0\left(\frac{p-u^2}{q-u^2}\right)$. β_2, β_4, β_6 satisfy $F_8=0$. We must take the + sign if (i) $F_8'=0$ and $K>0$ or (ii) $F_8''=0$ and $K<0$; otherwise the − sign.
(xx)	$p=2+(5\beta_2+4)^{\frac{1}{2}}$	The form of the equation is $y=y_0\left(\frac{p-u^2}{p+u^2}\right)$. β_2, β_4, β_6 satisfy $\beta_4=\frac{1}{7}[7\beta_2+24+4(3-\beta_2)\sqrt{5\beta_2+4}]$ and $\beta_6=\frac{1}{21}[(85\beta_2^2+16\beta_2-128)+16(3\beta_2-4)\sqrt{5\beta_2+4}]$. We find $p=\{2(3-\beta_2)\pm\lvert 3-\beta_2\rvert\sqrt{5\beta_2+4}\}/(3-\beta_2)$, but the only admissible value is $p=2+(5\beta_2+4)^{\frac{1}{2}}$.

At the point I we accordingly find the Type (viii) curve, viz.

$$y=y_0\,e^{-\frac{M}{P-u^2}},\quad \text{where}\quad u=\frac{x}{\sigma}.$$

We have managed to obtain limits to the expressions

$$\frac{9\beta_6(5\beta_2-9)-49\beta_4^2+210\beta_2\beta_4-125\beta_2^3}{4[5\beta_2^2-(6-\beta_2).\beta_4]},$$

$$\frac{9(3-\beta_2)\beta_6+7\beta_4^2-50\beta_2\beta_4+25\beta_2^3}{2[5\beta_2^2-(6-\beta_2)\beta_4]},$$

$$\frac{14\beta_4^2-9\beta_2\beta_6-5\beta_2^2\beta_4}{5\beta_2^2-(6-\beta_2).\beta_4},$$

when $\beta_4\to\dfrac{5\beta_2^2}{6-\beta_2}$ and $\beta_6\to\dfrac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2}$.

Note α. (1) P and M are imaginary for $\beta_2<1{\cdot}8$ and for such values of β_2 we do not get a fourth order Pearson curve at the point I.

(2) If $\beta_2=1{\cdot}8$, $M=0$ and we get $y=$ const.

(3) Let us see for what values of β_2, P and M are positive.

Now $$\sqrt{\frac{(\beta_2-1)(5\beta_2-9)}{6-\beta_2}} \gtrless 2,$$

according as $$(\beta_2-1)(5\beta_2-9) \gtrless 4(6-\beta_2),$$

" " $$5(\beta_2-1)(\beta_2-3) \gtrless 0,$$

" " $$\beta_2 \gtrless 3.$$

Hence P is always positive.

Again, $$3\sqrt{\beta_2-1} \gtrless \sqrt{(5\beta_2-9)(6-\beta_2)},$$

according as $$9(\beta_2-1) \gtrless (5\beta_2-9)(6-\beta_2),$$

" " $$5(\beta_2-3)^2 \gtrless 0.$$

Hence $3\sqrt{\beta_2-1} > \sqrt{(5\beta_2-9)(6-\beta_2)}$ for all values of β_2. Accordingly $M>0$ for all values of $\beta_2 > 1{\cdot}8$.

(4) When $\beta_2 = 3$, we must find P and M from the difference equation, which is easily found to be

$$(2s+1)P^2\mu_{2s} - 2P(2s+3)\frac{\mu_{2s+2}}{\mu_2} + (2s+5)\frac{\mu_{2s+4}}{\mu_2^2} = -2M\frac{\mu_{2s+2}}{\mu_2}.$$

Hence for $s=0$, $P^2 - 6P + 5\beta_2 = -2M$,

$s=1$, $3P^2 - 10P\beta_2 + 7\beta_4 = -2M\beta_2$.

Put $\left\{\begin{matrix}\beta_2 = 3\\ \beta_4 = 15\end{matrix}\right\}$ in the equations above and solve for P and M; therefore $P=5$ and $M=5$.

When $\beta_2 = 3$, $\beta_4 = 15$ and $\beta_6 = 91\frac{2}{3}$ we get the curve

$$y = y_0 e^{\frac{-5}{5-u^2}}.$$

The ordinates of this curve and the corresponding normal curve are plotted in Diagram II, showing how a change in β_6 changes the shape of the resulting frequency curve.

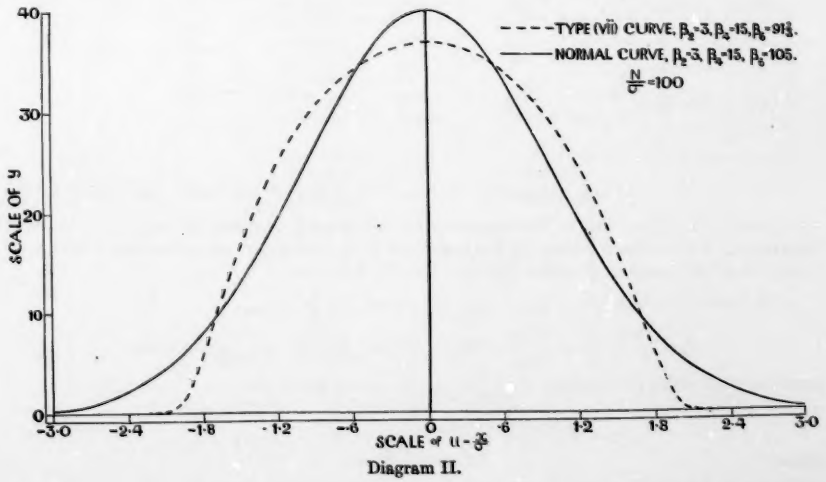


Diagram II.

The constants of integration for the transition types found by considering the areas under the curves are as follows:

TABLE VI. *Values of* y_0 (continuation of Table IV).

Type	Value for N	Expression for y_0
(vii)	$N\bar{\xi}=Sn_x\lvert x\rvert$	$\dfrac{N\bar{\xi}}{m\sigma^2 E_i\left(-\dfrac{m}{p}\right)+e^{-m/p}\left\{p\sigma^2-\dfrac{h^2}{12}-\dfrac{7h^4}{480}\dfrac{m}{p^2\sigma^2}\cdots\right\}}$ *
(ix)	$N=2y_0\sqrt{p}.\sigma\int_0^1\left(\dfrac{1-x^2}{1+x^2}\right)^n dx$	$\dfrac{N}{2\sigma\sqrt{p}}\dfrac{1}{I_n}$ †
(x)	$N=2y_0\,\sigma\sqrt{p}\int_0^1\left(\dfrac{1+x^2}{1-x^2}\right)^n dx$	$\dfrac{N}{2\sigma\sqrt{p}}\dfrac{1}{I_n}$
(xi)	$N=2y_0\int_0^{\sigma\sqrt{q}}\left(\dfrac{q\sigma^2-x^2}{x^2}\right)^{\frac{m}{q}}dx$	$\dfrac{N\sqrt{\pi}}{2\sigma\sqrt{q}}\dfrac{1}{\Gamma\left(\frac{1}{2}-l\right)\Gamma(1+l)}$ where $l=m/q$
(xii)	$N=2y_0\int_0^{\sigma\sqrt{q}}\left(\dfrac{x^2}{q\sigma^2-x^2}\right)^{\frac{m}{q}}dx$	$\dfrac{N\sqrt{\pi}}{2\sigma\sqrt{q}}\dfrac{1}{\Gamma\left(\frac{1}{2}+l\right)\Gamma(1-l)}$
(xiii)	$N=2y_0\int_0^{\sigma\sqrt{q}}\left(\dfrac{q\sigma^2-x^2}{p\sigma^2+x^2}\right)dx$	$\dfrac{N\sqrt{p}}{2\sigma\left[(q+p)\tan^{-1}\sqrt{q/p}-\sqrt{pq}\right]}$
(xiv)	$N=2y_0\int_0^{\sigma\sqrt{p}}\left(\dfrac{p\sigma^2-x^2}{p\sigma^2+x^2}\right)dx$	$\dfrac{N\sqrt{q}}{2\sigma\left[(p+q)\tan^{-1}\sqrt{p/q}-\sqrt{pq}\right]}$
(xv)	$N=2y_0\int_0^{\sigma\sqrt{p}}\left(\dfrac{p\sigma^2-x^2}{q\sigma^2-x^2}\right)dx$	$\dfrac{N\sqrt{q}}{\sigma\left[2\sqrt{pq}-q-p\left\{\log_e(\sqrt{q}+\sqrt{p})-\log_e(\sqrt{q}-\sqrt{p})\right\}\right]}$
(xx)	$N=2y_0\int_0^{\sigma\sqrt{p}}\left(\dfrac{p\sigma^2-x^2}{p\sigma^2+x^2}\right)dx$	$\dfrac{N}{2\sigma\sqrt{p}\left[\dfrac{\Pi}{2}-1\right]}$

* $E_i(-a)=\int_{\infty}^{a}\dfrac{e^{-z}}{z}dz$ and is tabled in *Mathematical Tables* of the British Association for the Advancement of Science, Vol. I. The expression is easily found by replacing the summation by one of Integration. Unfortunately I have not had time to try it out sufficiently, but in the trials I have made it gave satisfactory results. $\bar{\xi}$ denotes the mean absolute derivation.

† The reduction formula is $I_{n+1}=I_{n-1}-\dfrac{I_n}{n}$, $I_0=1$, $I_1=\dfrac{\pi}{2}-1$,

$$I_{\frac{1}{2}}=\frac{\sqrt{\pi}}{4}\left\{\frac{\Gamma\left(\frac{1}{4}\right)}{\Gamma\left(\frac{3}{4}\right)}-4\frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{1}{4}\right)}\right\}=\cdot71196,\quad I_{-\frac{1}{2}}=\frac{\sqrt{\pi}}{4}\left\{\frac{\Gamma\left(\frac{1}{4}\right)}{\Gamma\left(\frac{3}{4}\right)}+4\frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{1}{4}\right)}\right\}=1\cdot91010$$

and from these values I constructed Table VII (p. 151). I also proved that

$$I_{r+\theta}=I_r+\theta\Delta I_r-\frac{\theta(1-\theta)}{2!}\Delta^2 I_r+\ldots,$$

where $\Delta I_r=I_{r+1}-I_r$ etc.,

and hence we can find $I_{r+\theta}$ from I_r, I_{r+1} etc. by ordinary interpolation formulae.

Note β. The Types (xi) and (xii) are rather interesting. Consider Type (xi) which for convenience I shall write in the form

$$y = y_0 \left(\frac{b^2 - x^2}{x^2}\right)^l.$$

We have

$$N\mu_{2s} = 2\int_0^b y_0 x^{2s} \left(\frac{b^2 - x^2}{x^2}\right)^l dx$$

$$= y_0 b^{2s+1} B(s - l + \tfrac{1}{2}, l + 1) = y_0 b^{2s+1} \frac{\Gamma(s + \frac{1}{2} - l)\,\Gamma(l+1)}{\Gamma(s + \frac{3}{2})}.$$

TABLE VII. *Table of I_n and its differences.*

I_n	I	$\Delta^2 I_n$	$\Delta^4 I_n$	$\Delta^6 I_n$
		+	+	+
I_2	·429 2037	97229	13063	3070
$I_{2\cdot5}$	·387 8376	64951	7028	1389
I_3	·356 1945	45737	4063	673
$I_{3\cdot5}$	·331 0465	33551	2487	358
I_4	·310 4722	25428	1594	207
$I_{4\cdot5}$	·293 2529	19792	1059	116
I_5	·278 5765	15750	730	70
$I_{5\cdot5}$	·265 8792	12767	518	42
I_6	·254 7569	10513	375	36
$I_{6\cdot5}$	·244 9113	8778	274	19
I_7	·236 1170	7418	209	—
$I_{7\cdot5}$	·228 2005	6833	163	—
I_8	·221 0259	5456	125	—
$I_{8\cdot5}$	·214 4845	4741	95	—
I_9	·208 4887	4151	78	—
$I_{9\cdot5}$	·202 9670	3657	65	—
I_{10}	·197 8605	3240	53	—
I_{10}	·197 86048	116081	5687	593
I_{11}	·188 70268	93518	3970	365
I_{12}	·180 70569	76641	2847	233
I_{13}	·173 64387	63735	2088	153
I_{14}	·167 34846	53675	1563	103
I_{15}	·161 69041	45704	1190	72
I_{16}	·156 56911	39295	920	50
I_{17}	·151 90484	34077	722	32
I_{18}	·147 63353	29778	574	34
I_{19}	·143 70298	26202	458	10
I_{20}	·140 07021	23200	376	21

$I_{-\cdot5} = ·4310\,632$, $I_0 = 1·0$, $I_{+\cdot5} = ·711\,9587$,
$I_{1·0} = ·570\,7963$, $I_{1·5} = ·486\,1816$.

Put

$$s = 0, \quad N = by_0 \frac{\Gamma(\frac{1}{2} - l)\,\Gamma(1 + l)}{\Gamma(\frac{3}{2})},$$

$$s = 1, \quad \sigma^2 = b^2(\tfrac{1}{2} - l)/\tfrac{3}{2},$$

$$s = 2, \quad \beta_2 = \frac{3(3b^2 - 2m)}{5(b^2 - 2m)},$$

$$s = 3, \quad \beta_4 = \tfrac{1}{7}\beta_2(10\beta_2 - 3),$$

$$s = 4, \quad \beta_6 = \tfrac{1}{63}\beta_2(10\beta_2 - 3)(15\beta_2 - 6),$$

giving $$m = \frac{\sigma^2(5\beta_2 - 9)}{4}, \quad b^2 = \frac{\sigma^2(5\beta_2 - 3)}{2}, \quad l = \frac{5\beta_2 - 9}{2(5\beta_2 - 3)}.$$

Hence in every plane of section by a plane $\beta_2 =$ constant we get at one point of the curve of section of $F_0 = 0$, viz. when β_4 and β_6 have the values above a special Type (xi) curve with the constants expressible in terms of β_2 only. The difference equation is now

$$(2s+1)\,a^2\mu_{2s} - (2s+3)\frac{\mu_{2s+2}}{\mu_2} = 2m\mu_{2s}, \qquad (b^2 = a^2\sigma^2).$$

Put

$$s = 0, \qquad a^2 - 3 = 2m,$$

$$s = 1, \qquad 3a^2 - 5p_2 = 2m,$$

$$s = 2, \qquad 5a^2\beta_2 - 7\beta_4 = 2m\beta_2.$$

The first two of these equations give us the values found above, while the last two equations give us the general values.

II. The Transition Surfaces.

We have seen in Section I (*g*) (p. 143) how a knowledge of the curves of section of the transition surfaces for different values of β_2 will enable us to determine what type of frequency curve corresponds to any given values of β_2, β_4 and β_6.

In this Section we suppose β_2 known and then deduce general formulae in terms of β_2, to simplify the calculation of the β_4- and β_6-ordinates of the transition curves.

(*a*) The auxiliary co-ordinates r and Φ.

There is no way of finding the approximate range of values which β_2, β_4 and β_6 can assume. The value Professor Pearson finds for β_4 for his symmetrical curves is

$$\beta_4 = \frac{5\beta_2^2}{6 - \beta_2}.$$

The lowest possible value for β_4 is β_2^2, and I suggest taking as the range for β_4 the expression

$$2\left(\frac{5\beta_2^2}{6-\beta_2} - \beta_2^2\right) = \frac{2\beta_2^2(\beta_2 - 1)}{6 - \beta_2};$$

in other words, I assume that

$$\beta_2^2 \leq \beta_4 \leq \beta_2^2 + \frac{2\beta_2^2(\beta_2 - 1)}{6 - \beta_2}.$$

The only justification that I have for this assumption is that in all the symmetrical curves I have fitted I found β_4 well within the limits

$$\beta_2^2 \text{ and } \frac{\beta_2^2(\beta_2 + 4)}{(6 - \beta_2)}.$$

To make the model, I shall give β_2 different values and find the corresponding curves of section in the different β_2-planes. The curves of section depend then on β_4 and β_6 and it is an advantage to calculate the same number of β_6-ordinates for each curve, and at equal intervals of β_4, for each β_2. But the suggested range of possible values for β_4 varies as β_2 varies, e.g. the range is from 4 to 6 when $\beta_2 = 2$ and from 25 to 225 when $\beta_2 = 5$, and hence the interval between the successive values of β_4 must vary as β_2 varies. Similarly we must also have a proportionate unit for β_6.

All the surfaces pass through the point I, whose co-ordinates are

$$\beta_2 = \beta_2,$$

$$\beta_4 = \frac{5\beta_2^2}{6-\beta_2},$$

$$\beta_6 = \frac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2}.$$

I propose calculating the β_6-ordinates at intervals of

$$\beta_4 = \frac{1}{7}\frac{\beta_2^2(\beta_2-1)}{6-\beta_2} = \frac{1}{7}\left(\frac{5\beta_2^2}{6-\beta_2} - \beta_2^2\right)$$

and taking the unit for β_6 as

$$\frac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2} - \beta_2^3 = \frac{\beta_2^3(124-9\beta_2)(\beta_2-1)}{9(6-\beta_2)^2}.$$

In this way I introduce two auxiliary co-ordinates r and Φ, which I shall measure from the point I as origin. The relations between β_4 and r, β_6 and Φ, r and Φ_x then become

$$\beta_4 = \frac{5\beta_2^2}{6-\beta_2} + \frac{r}{7}\frac{\beta_2^2(\beta_2-1)}{6-\beta_2},$$

$$\Phi = \beta_6 - \frac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2} \Big/ \frac{\beta_2^3(124-9\beta_2)(\beta_2-1)}{9(6-\beta_2)^2}.$$

$$\Phi_x = \frac{\phi_x\left(\frac{5\beta_2^2}{6-\beta^2} + \frac{r}{7}\frac{\beta_2^2(\beta_2-1)}{6-\beta_2}\right) - \frac{25\beta_2^3(\beta_2+8)}{(6-\beta_2)^2}}{\frac{\beta_2^3(124-9\beta_2)(\beta_2-1)}{9(6-\beta_2)^2}},$$

where ϕ_x is the β_6-ordinate on the surface $F_x = 0$.

Let us represent graphically these relations.

Let OQ, OR represent the axis of β_4 and β_6,

IQ', IR' „ „ „ r and Φ,

O be the point (β_2^2, β_2^3),

I „ „ $\left(\frac{5\beta_2^2}{6-\beta_2}, \frac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2}\right)$.

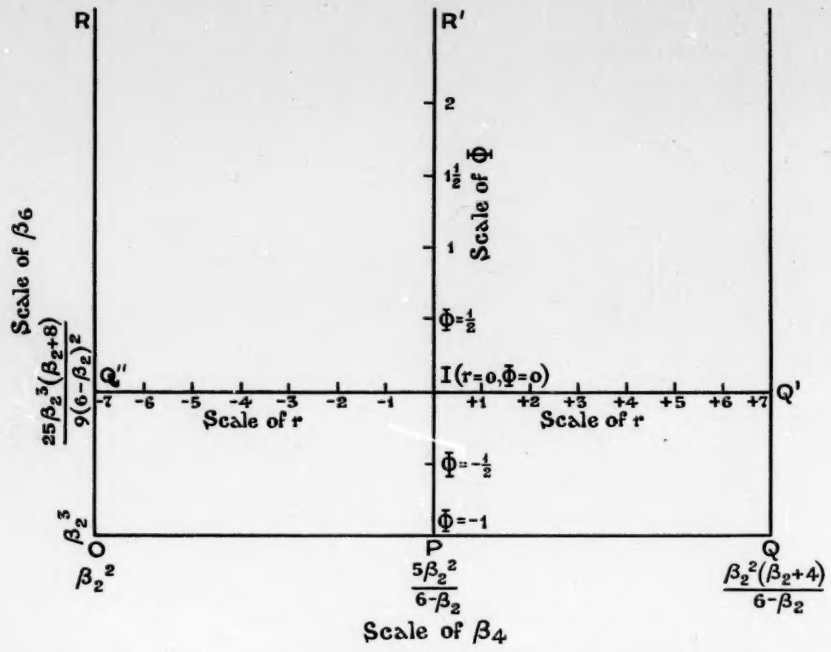


Diagram III.

Then
$$OP = \frac{\beta_2^2(\beta_2 - 1)}{6 - \beta_2},$$

$$PI = \beta_2^3(124 - 9\beta_2)(\beta_2 - 1)/9\,(6 - \beta_2)^2.$$

I calculate the β_6-ordinates at intervals of $\frac{1}{7}OP$, transfer to the point I as origin and divide by PI, to give me the corresponding value of Φ.

The following table gives values of PI and $\frac{1}{7}OP$ for different values of β_2.

TABLE VIII.

β_2	$\frac{1}{7}OP$	PI	β_2	$\frac{1}{7}OP$	PI
1·25	·01175	·27117	3·5	1·75	176·2639
1·5	·03571	1·02315	4	3·4286	469·3333
1·75	·07721	2·67660	4·5	6·75	1315·125
2	·14286	5·88889	5	14·2857	4388·889
2·5	·38265	21·57738	5·5	38·8922	24789·88
3	·85714	64·66667	6	∞	∞

Note that since $\beta_4 = \dfrac{5\beta_2^2}{6-\beta_2}$ is negative for $\beta_2 > 6$ we cannot use the r and Φ coordinates for $\beta_2 > 6$. I shall not make any diagrams for such values of β_2.

(*b*) Section I $f(\alpha)$ (p. 142) gives the relations

$$\phi_0 = \frac{14\beta_4^2 - 5\beta_2^2\beta_4}{9\beta_2} = \frac{f_0}{9\beta_2},$$

$$\phi_2 = \frac{-\{7\beta_4^2 - 50\beta_2\beta_4 + 25\beta_2^3\}}{9(3-\beta_2)} = \frac{f_2}{9(3-\beta_2)} \qquad (\beta_2 \neq 3),$$

$$\phi_4 = \frac{49\beta_4^2 - 210\beta_2\beta_4 + 125\beta_2^3}{9(5\beta_2 - 9)} = \frac{f_4}{9(5\beta_2-9)} \qquad (\beta_2 \neq 1{\cdot}8),$$

$$\phi_6', '' = \frac{b \pm 4|K|\sqrt{L}}{9(3-\beta_2)^2} \qquad (\beta_2 \neq 3),$$

$$\phi_8', '' = \frac{-\beta \pm 4|K|\sqrt{Q}}{27(\beta_2-1)(7\beta_2-15)} \qquad (\beta_2 \neq 2\tfrac{1}{7}),$$

where

$$b = (3-\beta_2)f_2 - 8\beta_2 . K,$$

$$B = (5\beta_2 - 9)f_4 - 4b,$$

$$K = 5\beta_2^2 - \overline{6-\beta_2} . \beta_4,$$

$$L = \beta_2^2(19 - 5\beta_2) - 7(3-\beta_2)\beta_4,$$

$$Q = 49\beta_4^2 - 14\beta_4(37\beta_2 - 30) + 7\beta_2^2(60\beta_2 - 53).$$

I assume for the present $\beta_2 \neq 1{\cdot}8$, $\beta_2 \neq 2\frac{1}{7}$, $\beta_2 \neq 3$.

Write

$$l = \frac{5\beta_2^2}{6-\beta_2},$$

$$m = \frac{r}{7}\frac{\beta_2^2(\beta_2-1)}{(6-\beta_2)},$$

$$n = \frac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2}.$$

We want the values of $\phi_x(l+m)$, or $f_x(l+m)$. Using Taylor's theorem, we have

$$f_x(l+m) = f_x(l) + mf_x'(l) + \frac{m^2}{2!}f_x''(l) + \ldots$$

for $x = 0, 2, 4, 6, 8$. The expressions f_0, f_2, f_4, L, and Q are quadratic expressions in β_4 and hence all their third and higher order differentials with respect to β_4 vanish. Again, we simplify the algebra by using the result

$$\phi_x(l) = \frac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2} = n$$

for $x = 0, 2, 4, 6, 8$ since all the surfaces pass through the point I.

Now
$$f_0'(\beta_4) = 28\beta_4 - 5\beta_2^2,$$
$$f_0''(\beta_4) = 28,$$
$$f_0'(l) = \frac{5\beta_2^2}{6-\beta_2}(\beta_2+22),$$
$$f_0(l) = 9\beta_2\phi_0(l) = 9\beta_2 . n.$$

Hence
$$f_0(l+m) = f_0(l) + mf_0'(l) + \frac{m^2}{2!}f_0''(l)$$
$$= 9\beta_2 . n + \frac{r}{7}\frac{\beta_2^2(\beta_2-1)}{6-\beta_2}\frac{5\beta_2^2(\beta_2+22)}{6-\beta_2} + \frac{1}{2}\frac{r^2}{49}\frac{\beta_2^4(\beta_2-1)^2}{(6-\beta_2)^2}.$$

But
$$\Phi_0 = \frac{\phi_0(l+m)-n}{\frac{\beta_2^3(124-9\beta_2)(\beta_2-1)}{9(6-\beta_2)^2}}$$
$$= \frac{r\{2r(\beta_2-1)+5(\beta_2+22)\}}{7(124-9\beta_2)}.$$

Similarly, after lengthy algebra,

$$\Phi_2 = \frac{r\{r\beta_2(\beta_2-1)+60(2\beta_2-5)\}}{7(124-9\beta_2)(\beta_2-3)},$$
$$\Phi_4 = \frac{r\{r\beta_2(\beta_2-1)+20(5\beta_2-9)\}}{(5\beta_2-9)(124-9\beta_2)},$$
$$\Phi_6','' = \frac{r\{r\beta_2(\beta_2-1)(\beta_2-3)+4\psi\}}{7(\beta_2-3)^2(124-9\beta_2)};$$

where
$$\psi = 32\beta_2^2 - 189\beta_2 + 297 \mp \{(6-\beta_2)^3(\beta_2-1)[(5\beta_2-9)+r(\beta_2-3)]\}^{\frac{1}{2}},$$
$$\Phi_8','' = \frac{r\{r\beta_2(31\beta_2-51)+4\chi\}}{21(124-9\beta_2)(7\beta_2-15)};$$

where
$$\chi = 9(83\beta_2-183) \mp (6-\beta_2)[84(5\beta_2-9)+72r(2\beta_2-5)+\beta_2^2r^2]^{\frac{1}{2}}.$$

...(12)

Note $\Phi_6' > \Phi_6''$, $\Phi_8' > \Phi_8''$ according as $\beta_2 \gtrless \frac{15}{7}$. By giving β_2 its constant value and putting $r = -7$ to $r = +7$ we can easily calculate the ordinates.

Not only have I solved the difficulty of choosing a suitable scale for β_4 and β_6 and so made the construction of a model practicable, but I have simplified the calculation of the ordinates of the transition surfaces considerably.

(c) *Calculation of the Ordinates.*

(i) Substituting for β_2 in Equations (12) we find the following relations between Φ_0, Φ_2, Φ_4, and r for different values of β_2:

TABLE IX. *Values of* $\Phi_0, \ldots \Phi_4$ *for values of* β_2.

β_2	Φ_0	Φ_2	Φ_4
1	$\frac{r}{7}$	$\frac{18r}{161}$	$\frac{4r}{23}$
1·25	$\frac{r[465+2r]}{3157}$	$\frac{r[2400-5r]}{22099}$	$\frac{r[880-5r]}{4961}$
1·5	$\frac{r[235+2r]}{1547}$	$\frac{r[160-r]}{1547}$	$\frac{r[40-r]}{221}$
1·75	$\frac{r[475+6r]}{3031}$	$\frac{r[1440-21r]}{15155}$	$\frac{r[80-21r]}{433}$
2	$\frac{r[60+r]}{371}$	$\frac{r[30-r]}{371}$	$\frac{r[10+r]}{53}$
2·5	$\frac{r[245+6r]}{1421}$	$\frac{-15r^2}{1421}$	$\frac{r[280+15r]}{1421}$
3	$\frac{r[125+4r]}{679}$	———	$\frac{r[20+r]}{97}$
3·5	$\frac{r[51+2r]}{259}$	$\frac{r[96+7r]}{259}$	$\frac{r[136+7r]}{629}$
4	$\frac{r[45+3r]}{154}$	$\frac{r[65+3r]}{308}$	$\frac{r[55+3r]}{242}$
4·5	$\frac{r[265+14r]}{1169}$	$\frac{r[320+21r]}{1169}$	$\frac{r[120+7r]}{501}$
5	$\frac{r[135+8r]}{553}$	$\frac{r[150+10r]}{553}$	$\frac{r[80+5r]}{316}$
5·5	$\frac{r[275+18r]}{1043}$	$\frac{r[1440+99r]}{5215}$	$\frac{r[1480+99r]}{5513}$

In the same way we find the following table, which gives the relations between Φ_6, Φ_8 and r for different values of β_2:

TABLE X. *Values of* Φ_6 *and* Φ_8 *for Values of* β_2.

β_2	Φ_6	Φ_8
1	$\frac{4r}{23}$	$\frac{r\{3600+20r\pm20[r^2-216r-336]^{\frac{1}{2}}\}}{19320}$
1·25	$\frac{r[28352-35r\mp152(-19.7r+11)^{\frac{1}{2}}]}{154693}$	$\frac{r[45648+245r\pm76\{25r^2-2880r-3696\}^{\frac{1}{2}}]}{236775}$
1·5	$\frac{r[304-r\mp12(-6.\overline{r+1})^{\frac{1}{2}}]}{1547}$	$\frac{r[312+r\pm4(r.\overline{r-64}-56)^{\frac{1}{2}}]}{1547}$
1·75	$\frac{r[16448-105r\mp136\{51(-5r-1)\}^{\frac{1}{2}}]}{75775}$	$\frac{r[21744-91r\pm68\{49r^2-1728r-336\}^{\frac{1}{2}}]}{100023}$
2	$\frac{r[94-r\mp(256.\overline{1-r})^{\frac{1}{2}}]}{371}$	$\frac{r[306-11r\pm8(84-r.\overline{72-4r})^{\frac{1}{2}}]}{1113}$
2·5	$\frac{r[784-15r\mp(42.\overline{7-r})^{\frac{1}{2}}]}{1421}$	$\frac{r[3528+265r\mp28(1176+25r^2)^{\frac{1}{2}}]}{21315}$
3	$\frac{r[600+r(36+r)]}{2716}$	$\frac{r[132+7r\mp2(56+r.\overline{8+r})^{\frac{1}{2}}]}{2716}$
3·5	$\frac{r[176+7r\mp20(2.\overline{17+r})^{\frac{1}{2}}]}{259}$	$\frac{r[3096+161r\mp4(2856+r.\overline{576+49r})^{\frac{1}{2}}]}{14763}$
4	$\frac{r[53+3r\mp(24.\overline{11+r})^{\frac{1}{2}}]}{154}$	$\frac{r[1341+73r\mp2(924+r.\overline{216+16r})^{\frac{1}{2}}]}{6006}$
4·5	$\frac{r[336+21r\mp4(14.\overline{9+r})^{\frac{1}{2}}]}{1169}$	$\frac{r[3048+177r\mp4(504+r.\overline{128+9r})^{\frac{1}{2}}]}{12859}$
5	$\frac{r[152+10r\mp(64+8r)^{\frac{1}{2}}]}{553}$	$\frac{r[2088+130r\mp(1344+r.\overline{360+25r})^{\frac{1}{2}}]}{8295}$
5·5	$\frac{r[7216+495r\mp12(74+10r)^{\frac{1}{2}}]}{26075}$	$\frac{r[39384+2629r\mp4(6216+r.\overline{1728+121r})^{\frac{1}{2}}]}{147063}$

(*d*) From the previous tables it is now an easy matter to calculate the Φ-ordinates of each of the transition surfaces for integral values of r from -7 to $+7$ and β_2 from $1{\cdot}25-({\cdot}25)-2{\cdot}0$ and $2{\cdot}0-({\cdot}5)-5{\cdot}5$. The results are given in Tables XI—XXII and the ordinates are plotted in Diagrams VI—XIX.

TABLE XI. *The Ordinates of the Transition Curves for* $\beta_2 = 1{\cdot}25$.

r	Φ_0	Φ_2	Φ_4	Φ_6'	Φ_6''	Φ_8'	Φ_8''
−7	−1·00000	−·77130	−1·29107	−1·10922	−1·47886	−1·59766	−1·00000
−6	−·86094	−·65976	−1·10058	−·96474	−1·25090	−1·35127	−·88771
−5	−·72062	−·54867	−·91211	−·81714	−1·02696	−1·10891	−·76726
−4	−·57903	−·43803	−·72566	−·66610	−·80737	−·87104	−·63817
−3	−·43617	−·32784	−·54122	−·51124	−·59250	−·63829	−·49983
−2	−·29205	−·21811	−·35880	−·35263	−·38230	−·41131	−·35158
−1	−·14666	−·10883	−·17839	—	—	—	—
+1	+·14793	+·10838	+·17638	—	—	—	—
+2	+·29712	+·21630	+·35074	—	—	—	—
+3	+·44758	+·32377	+·52308	—	—	—	—
+4	+·59930	+·43079	+·69341	—	—	—	—
+5	+·75230	+·53735	+·86172	—	—	—	—
+6	+·90656	+·64347	+1·02802	—	—	—	—
+7	+1·06208	+·74913	+1·19230	—	—	—	—

TABLE XII. *The Ordinates of the Transition Curves for* $\beta_2 = 1{\cdot}5$.

r	Φ_0	Φ_2	Φ_4	Φ_6'	Φ_6''	Φ_8'	Φ_8''
−7	−1·00000	−·75566	−1·48869	−1·08145	...	...	−1·00000
−6	−·86490	−·64383	−1·24887	−·94741	...	...	−·89083
−5	−·72721	−·53329	−1·01810	−·80870	−1·18871	−1·21202	−·77246
−4	−·58694	−·42405	−·79638	−·66474	−·92802	−·94839	−·64438
−3	−·44086	−·31610	−·58371	−·51473	−·67596	−·69263	−·50582
−2	−·29864	−·20944	−·38009	−·35760	−·43361	−·44586	−·35569
−1	−·15061	−·10407	−·18552	−·19716	−·19716	−·20879	−·19328
+1	+·15320	+·10278	+·17647	—	—	—	—
+2	+·30899	+·20427	+·34389	—	—	—	—
+3	+·46736	+·30446	+·50226	—	—	—	—
+4	+·62831	+·40336	+·65158	—	—	—	—
+5	+·79186	+·50097	+·79186	—	—	—	—
+6	+·95798	+·59729	+·92308	—	—	—	—
+7	+1·12670	+·69231	+1·04525	—	—	—	—

— Denotes that the value of Φ is imaginary.
... Denotes that Φ is now a large negative quantity.

TABLE XIII. *The Ordinates of the Transition Curves for* $\beta_2 = 1{\cdot}75$.

r	Φ_0	Φ_2	Φ_4	Φ_6'	Φ_6''	Φ_8'	Φ_8''
−7	−1·00000	−·73303	...	−1·06418	...	...	−1·00000
−6	− ·86902	−·61999	...	− ·93813	−1·76641	−1·78012	− ·89407
−5	− ·73408	−·50973	−2·13626	− ·80600	−1·43392	−1·44152	− ·77788
−4	− ·59518	−·40224	−1·51501	− ·66695	−1·11390	−1·11741	− ·65082
−3	− ·45233	−·29753	− ·99076	− ·51979	− ·80754	− ·80869	− ·51203
−2	− ·30551	−·19558	− ·56351	− ·36277	− ·51657	− ·51672	− ·36012
−1	− ·15473	−·09640	− ·23326	− ·19281	− ·24408	− ·24405	− ·19255
+1	+ ·15869	+·09363	+ ·13626	—	—	—	—
+2	+ ·32135	+·18449	+ ·17552	—	—	—	—
+3	+ ·48796	+·27258	+ ·11778	—	—	—	—
+4	+ ·65853	+·35790	− ·03695	—	—	—	—
+5	+ ·83306	+·44045	− ·28868	—	—	—	—
+6	+1·01155	+·52022	− ·63741	—	—	—	—
+7	+1·19400	+.59723	−1·08314	—	—	—	—

TABLE XIV. *The Ordinates of the Transition Curves for* $\beta_2 = 2$.

r	Φ_0	Φ_2	Φ_4	Φ_6'	Φ_6''	Φ_8'	Φ_8''
−7	−1·00000	−·69811	− ·39623	−1·07066	...	...	−1·00000
−6	− ·87332	−·58221	− ·45283	− ·94881	...	...	− ·89745
−5	− ·74124	−·47170	− ·47170	− ·81952	...	...	− ·78351
−4	− ·60377	−·36658	− ·45283	− ·67087	−1·44234	−1·85820	− ·65752
−3	− ·46092	−·26685	− ·39623	− ·52561	−1·04313	−1·30901	− ·51848
−2	− ·31267	−·17251	− ·30189	− ·36812	− ·66692	− ·81395	− ·36484
−1	− ·15903	−·08356	− ·16981	− ·19507	− ·31706	− ·37574	− ·19390
+1	+ ·16442	+·07817	+ ·20755	+ ·25067	+ ·25067	+ ·23630	+ ·29380
+2	+ ·33423	+·15094	+ ·45283	—	—	—	—
+3	+ ·50943	+·21833	+ ·73585	—	—	—	—
+4	+ ·69003	+·28032	+1·05660	—	—	—	—
+5	+ ·87601	+·33693	+1·41509	—	—	—	—
+6	+1·06739	+·38814	+1·81132	—	—	—	—
+7	+1·26415	+·43396	+2·24528	—	—	—	—

TABLE XV. *The Ordinates of the Transition Curves for* $\beta_2 = 2{\cdot}5$.

r	Φ_0	Φ_2	Φ_4	Φ_6'	Φ_6''	Φ_8'	Φ_8''
−7	−1·00000	−·51724	− ·86207	−1·03466	...	− ·09885	−1·00000
−6	− ·88248	−·38001	− ·80225	− ·92780	...	− ·18641	− ·90465
−5	− ·75651	−·26390	− ·72132	− ·81070	...	− ·23803	− ·79551
−4	− ·62210	−·16890	− ·61928	− ·68167	...	− ·25455	− ·67175
−3	− ·47924	−·09500	− ·49613	− ·53871	...	− ·23715	− ·53217
−2	− ·32794	−·04222	− ·35186	− ·37948	−1·91187	− ·18746	− ·37515
−1	− ·16819	−·01056	− ·18649	− ·20109	− ·92347	− ·10756	− ·19861
+1	+ ·17664	−·01056	+ ·20760	+ ·85397	+ ·22837	+ ·22347	+ ·13243
+2	+ ·36172	−·04222	+ ·43631	+1·63231	+ ·49014	+ ·47461	+ ·28692
+3	+ ·55524	−·09500	+ ·68614	+2·32636	+ ·79397	+ ·75595	+ ·46094
+4	+ ·75721	−·16890	+ ·95707	+2·92273	+1·15327	+1·06959	+ ·65239
+5	+ ·96763	−·26390	+1·24912	+3·39769	+1·59175	+1·41714	+ ·85966
+6	+1·18649	−·38001	+1·56228	+3·69653	+2·16413	+1·79979	+1·08156
+7	+1·41379	−·51724	+1·89655	+3·34483	+3·34483	+2·21839	+1·31724

— Denotes that the value of Φ is imaginary.
... Denotes that Φ is now a large negative quantity.

TABLE XVI. *The Ordinates of the Transition Curves for* $\beta_2 = 3$.

r	Φ_0	Φ_2*	Φ_4	Φ_6'*	Φ_6''	Φ_8'	Φ_8''
−7	−1·000 000		− ·938 144	−1·023 196		− ·711 340	−1·000 000
−6	− ·892 489		− ·865 979	− ·927 835		− ·678 057	− ·912 517
−5	− ·773 196		− ·773 196	− ·819 219		− ·619 984	− ·808 588
−4	− ·642 121	∞	− ·659 794	− ·695 140	∞	− ·538 150	− ·687 182
−3	− ·499 264		− ·525 773	− ·553 387		− ·433 846	− ·547 008
−2	− ·344 624		− ·371 134	− ·391 753		− ·308 493	− ·386 647
−1	− ·178 203		− ·195 876	− ·208 027		− ·163 476	− ·204 713
+1	+ ·189 985		+ ·216 495	+ ·234 536		+ ·228 460	+ ·180 965
+2	+ ·391 753		+ ·453 608	+ ·497 791		+ ·481 401	+ ·378 687
+3	+ ·605 302		+ ·711 340	+ ·791 973		+ ·759 358	+ ·592 630
+4	+ ·830 633	∞	+ ·989 691	+1·119 293	∞	+1·062 716	+ ·822 409
+5	+1·067 747		+1·288 660	+1·481 959		+1·391 753	+1·067 747
+6	+1·316 642		+1·608 247	+1·882 180		+1·746 666	+1·328 445
+7	+1·577 320		+1·948 454	+2·322 165		+2·127 600	+1·604 359

TABLE XVII. *The Ordinates of the Transition Curves for* $\beta_2 = 3{\cdot}5$.

r	Φ_0	Φ_2	Φ_4	Φ_6'	Φ_6''	Φ_8'	Φ_8''
−7	−1·000 000	−1·270 270	− ·968 203	−1·015 062	—	− ·867 236	−1·000 000
−6	− ·903 475	−1·250 975	− ·896 661	− ·931 081	—	− ·811 839	− ·919 516
−5	− ·791 506	−1·177 606	− ·802 862	− ·830 510	—	− ·728 977	− ·822 875
−4	− ·664 093	−1·050 193	− ·686 804	− ·710 728	—	− ·624 750	− ·703 978
−3	− ·521 236	− ·868 726	− ·548 490	− ·569 536	—	− ·498 792	− ·563 187
−2	− ·362 934	− ·633 205	− ·387 917	− ·405 062	—	− ·352 184	− ·399 425
−1	− ·189 189	− ·343 629	− ·205 087	− ·215 687	−1·089 332	− ·186 003	− ·211 613
+1	+ ·204 633	+ ·397 683	+ ·227 345	+1·169 884	+ ·243 243	+ ·236 605	+ ·204 633
+2	+ ·424 710	+ ·849 421	+ ·476 948	+2·419 215	+ ·515 148	+ ·498 185	+ ·427 914
+3	+ ·660 232	+1·355 212	+ ·748 808	+3·747 001	+ ·816 705	+ ·784 911	+ ·669 671
+4	+ ·911 197	+1·915 058	+1·042 925	+5·152 352	+1·148 807	+1·095 818	+ ·930 871
+5	+1·177 606	+2·528 958	+1·359 300	−·−·−	+1·512 259	+1·434 238	+1·208 179
+6	+1·459 459	+3·196 911	+1·697 933	−·−·−	+1·907 801	+1·796 979	+1·504 789
+7	+1·756 757	+3·918 919	+2·058 824	−·−·−	+2·336 106	+2·185 167	+1·819 574

TABLE XVIII. *The Ordinates of the Transition Curves for* $\beta_2 = 4$.

r	Φ_0	Φ_2	Φ_4	Φ_6'	Φ_6''	Φ_8'	Φ_8''
−7	−1·000 000	... ...	− ·983 471	−1·009 184	... ...	− ·934 732	−1·000 000
−6	− ·915 584	−1·051 948	− ·917 355	− ·936 840	... ...	− ·873 561	− ·930 635
−5	− ·811 688	− ·974 026	− ·826 446	− ·844 156	... ...	− ·786 513	− ·838 529
−4	− ·688 312	− ·857 143	− ·710 744	− ·728 273	... ...	− ·674 957	− ·722 313
−3	− ·545 454	− ·701 299	− ·570 248	− ·587 213	−1·127 073	− ·539 966	− ·580 913
−2	− ·383 117	− ·506 494	− ·404 959	− ·419 520	− ·801 259	− ·382 231	− ·413 640
−1	− ·201 299	− ·272 727	− ·214 876	− ·224 079	− ·425 272	− ·202 162	− ·220 082
+1	+ ·220 779	+ ·311 688	+ ·239 669	+ ·473 835	+ ·253 438	+ ·246 753	+ ·224 109
+2	+ ·461 039	+ ·662 338	+ ·504 132	+ ·995 630	+ ·536 838	+ ·520 268	+ ·470 075
+3	+ ·720 779	+1·051 948	+ ·793 388	+1·564 876	+ ·850 708	+ ·820 604	+ ·737 838
+4	+1·000 000	+1·480 519	+1·107 438	+2·181 134	+1·195 489	+1·146 468	+1·028 691
+5	+1·298 701	+1·948 052	+1·446 281	+2·844 024	+1·571 561	+1·500 218	+1·340 276
+6	+1·616 883	+2·454 545	+1·809 917	+3·553 208	+1·979 259	+1·882 872	+1·671 574
+7	+1·954 545	+3·000 000	+2·198 347	+4·308 392	+2·418 881	+2·290 781	+2·026 236

* See (f) (iv) on p. 164.

— Denotes that the value of Φ is imaginary.

−·−·− Denotes that Φ is a large positive quantity.

... ... Denotes that Φ is a large negative quantity.

TABLE XIX. *The Ordinates of the Transition Curves for* $\beta_2 = 4{\cdot}5$.

r	Φ_0	Φ_2	Φ_4	Φ_6'	Φ_6''	Φ_8'	Φ_8''
−7	−1·000 000	−1·035 928	− ·992 016	−1·004 994	−1·258 479	− ·969 516	−1·000 000
−6	− ·928 999	− ·995 723	− ·934 132	− ·944 792	−1·210 896	− ·912 209	− ·941 123
−5	− ·834 046	− ·919 589	− ·848 303	− ·859 995	−1·116 053	− ·826 372	− ·855 718
−4	− ·715 141	− ·807 528	− ·734 531	− ·747 763	− ·976 788	− ·713 384	− ·742 405
−3	− ·572 284	− ·659 538	− ·592 814	− ·606 517	− ·794 681	− ·573 985	− ·600 446
−2	− ·405 475	− ·475 620	− ·423 154	− ·435 247	− ·570 741	− ·408 522	− ·429 490
−1	− ·214 713	− ·255 774	− ·225 549	− ·233 249	− ·305 673	− ·217 164	− ·229 371
+1	+ ·238 666	+ ·291 702	+ ·253 493	+ ·345 876	+ ·264 903	+ ·258 673	+ ·242 922
+2	+ ·501 283	+ ·619 333	+ ·534 930	+ ·731 632	+ ·561 782	+ ·546 676	+ ·511 571
+3	+ ·787 853	+ ·982 891	+ ·844 311	+1·157 004	+ ·890 900	+ ·864 029	+ ·805 930
+4	+1·098 375	+1·382 378	+1·181 637	+1·621 772	+1·252 479	+1·209 498	+1·127 231
+5	+1·432 849	+1·817 793	+1·546 906	+2·125 749	+1·646 707	+1·586 826	+1·471 732
+6	+1·791 275	+2·289 136	+1·940 120	+2·668 771	+2·073 744	+1·992 285	+1·843 161
+7	+2·173 653	+2·796 407	+2·361 277	+3·250 698	+2·533 733	+2·427 124	+2·240 268

TABLE XX. *The Ordinates of the Transition Curves for* $\beta_2 = 5$.

r	Φ_0	Φ_2	Φ_4	Φ_6'	Φ_6''	Φ_8'	Φ_8''
−7	−1·000 000	−1·012 658	− ·996 835	−1·002 172	−1·073 778	− ·988 186	−1·000 000
−6	− ·943 942	− ·976 492	− ·949 367	− ·954 792	−1·041 591	− ·939 483	− ·952 742
−5	− ·858 951	− ·904 159	− ·870 253	− ·877 948	− ·966 537	− ·858 951	− ·874 623
−4	− ·745 027	− ·795 660	− ·759 494	− ·769 209	− ·851 044	− ·747 710	− ·764 526
−3	− ·602 170	− ·650 995	− ·617 089	− ·627 534	− ·696 155	− ·606 107	− ·622 102
−2	− ·430 380	− ·470 163	− ·443 038	− ·452 339	− ·502 453	− ·434 260	− ·447 235
−1	− ·229 656	− ·253 165	− ·237 342	− ·243 249	− ·270 313	− ·232 216	− ·239 875
+1	+ ·258 590	+ ·289 331	+ ·268 987	+ ·308 292	+ ·277 603	+ ·272 403	+ ·262 378
+2	+ ·546 112	+ ·614 828	+ ·569 620	+ ·654 410	+ ·589 713	+ ·577 340	+ ·554 908
+3	+ ·862 568	+ ·976 492	+ ·901 899	+1·038 232	+ ·936 451	+ ·914 455	+ ·877 950
+4	+1·207 957	+1·374 322	+1·265 823	+1·459 660	+1·317 917	+1·284 835	+1·230 414
+5	+1·582 278	+1·808 318	+1·661 392	+1·918 608	+1·734 195	+1·687 397	+1·613 386
+6	+1·985 533	+2·278 481	+2·088 608	+2·415 005	+2·185 356	+2·122 505	+2·026 501
+7	+2·417 722	+2·784 810	+2·547 468	+2·948 791	+2·671 463	+2·590 158	+2·469 758

TABLE XXI. *The Ordinates of the Transition Curves for* $\beta_2 = 5{\cdot}5$.

r	Φ_0	Φ_2	Φ_4	Φ_6'	Φ_6''	Φ_8'	Φ_8''
−7	−1·000 000	−1·002 685	− ·999 274	−1·000 537	−1·013 423	− ·997 335	−1·000 000
−6	− ·960 690	− ·973 346	− ·964 266	− ·966 696	− ·987 360	− ·960 930	− ·965 592
−5	− ·886 865	− ·906 040	− ·893 343	− ·897 836	− ·920 381	− ·888 767	− ·895 434
−4	− ·778 523	− ·800 767	− ·786 505	− ·792 488	− ·813 955	− ·781 465	− ·788 910
−3	− ·635 666	− ·657 526	− ·643 751	− ·650 209	− ·668 525	− ·638 844	− ·646 197
−2	− ·458 293	− ·476 318	− ·465 083	− ·470 782	− ·484 309	− ·461 002	− ·467 199
−1	− ·246 405	− ·257 143	− ·250 499	− ·254 075	− ·261 438	− ·248 080	− ·251 774
+1	+ ·280 920	+ ·295 110	+ ·286 414	+ ·299 942	+ ·291 506	+ ·288 123	+ ·283 238
+2	+ ·596 357	+ ·628 188	+ ·608 743	+ ·638 339	+ ·620 491	+ ·612 596	+ ·601 632
+3	+ ·946 309	+ ·999 233	+ ·966 987	+1·015 154	+ ·986 994	+ ·973 420	+ ·955 182
+4	+1·330 777	+1·408 245	+1·361 146	+1·430 355	+1·391 045	+1·370 595	+1·343 889
+5	+1·749 760	+1·855 225	+1·791 221	+1·883 917	+1·832 670	+1·804 121	+1·767 750
+6	+2·203 260	+2·340 173	+2·257 210	+2·375 818	+2·311 890	+2·273 998	+2·226 767
+7	+2·691 275	+2·863 087	+2·759 115	+2·906 040	+2·828 725	+2·780 227	+2·720 940

(*e*) The limiting curve $\beta_6 = \beta_2\beta_4$; the Φ co-ordinate of the second order Pearson curve.

(i) I have calculated ordinates at values of r from -7 to $+7$ and found the corresponding Φ-ordinates for different values of β_2. All these values of Φ will not be possible because β_2, β_4, and β_6 must satisfy the inequalities $\beta_4 > \beta_2^2$, $\beta_6 > \beta_2\beta_4 > \beta_2^3$. For all values of r from -7 to $+7$, $\beta_4 > \beta_2^2$ and $\beta_6 > \beta_2^3$. Let us consider now the surface $\beta_6 = \beta_2\beta_4$ and express this relation as an equation in r and Φ.

$$\Phi = \left\{\beta_6 - \frac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2}\right\} \Big/ \left\{\frac{\beta_2^3(124-9\beta_2)(\beta_2-1)}{9(6-\beta_2)^2}\right\}$$

$$= \frac{\beta_2\left\{\dfrac{5\beta_2^2}{6-\beta_2} + \dfrac{r}{7}\dfrac{\beta_2^2(\beta_2-1)}{6-\beta_2}\right\} - \dfrac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2}}{\dfrac{\beta_2^3(124-9\beta_2)(\beta_2-1)}{9(6-\beta_2)^2}},$$

since $$\beta_4 = \frac{5\beta_2^2}{6-\beta_2} + \frac{r}{7}\frac{\beta_2^2(\beta_2-1)}{6-\beta_2}.$$

Hence, on simplifying we find

$$\Phi = \frac{9r(6-\beta_2) - 490}{7(124-9\beta_2)} \qquad \text{.....(13).}$$

Let us denote this limiting surface $\beta_6 = \beta_2\beta_4$ by $F_1 = 0$ and this limiting value of Φ by Φ_1. For given β_2, Φ lies along a straight line which passes through the point ($r = -7$, $\Phi = -1$). I shall accordingly calculate Φ_1 for $r = 0$ and 7, and for different values of β_2'.

TABLE XXII *a*.

β_2	$\Phi_1(0)$	$\Phi_1(7)$	β_2	$\Phi_1(0)$	$\Phi_1(7)$
1·25	−·621	−·242	3·5	−·757	−·514
1·5	−·634	−·267	4	−·796	−·591
1·75	−·647	−·293	4·5	−·838	−·677
2	−·660	−·321	5	−·886	−·772
2·5	−·690	−·379	5·5	−·940	−·879
3	−·722	−·443			

(ii) The values of β_4 and β_6 corresponding to the second order symmetrical Pearson curves are

$$\beta_4 = \frac{5\beta_2^2}{6-\beta_2}, \text{ i.e. } r = 0 \text{ and } \beta_6 = \frac{35\beta_2^3}{(9-2\beta_2)(6-\beta_2)}.$$

Hence
$$\left.\begin{aligned}\Phi &= \left\{\frac{35\beta_2^3}{(9-2\beta_2)(6-\beta_2)} - \frac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2}\right\} \Big/ \frac{\beta_2^3(124-9\beta_2)(\beta_2-1)}{9(6-\beta_2)^2} \\ &= \frac{10(5\beta_2-9)}{(124-9\beta_2)(9-2\beta_2)}\end{aligned}\right\} \text{(14).}$$

Note that Φ is negative for $\beta_2 < 1{\cdot}8$ and for $\beta_2 > 4{\cdot}5$.

If $\beta_2 = 4{\cdot}5$, Φ becomes infinite. By substituting for β_2 we can find the Φ ordinate corresponding to that value of β_2. We find

TABLE XXII *b*.

β_2	Φ	β_2	Φ
1·25	−·038	2·5	+·086
1·5	−·023	3	+·206
1·75	−·002	3·5	+·483
2	+·019	4	+1·189

(*f*) I will now discuss the curves of section corresponding to critical values of β_2, viz. $\beta_2 = 1,\ 1{\cdot}8,\ 2\frac{1}{7},\ 3$.

(i) $\beta_2 = 1$. We now have the "two-lumps" distribution for which $\beta_2 = \beta_4 = \beta_6 = 1$. Now the points O and I coincide, in other words, for all values of r, β_4 always remains unity. I have added the curves of section in the case $\beta_2 = 1$ to each of the contour diagrams for the sake of completeness.

(ii) $\beta_2 = 1{\cdot}8$. The curve of section of the surface $F_4 = 0$ reduces to the two coincident straight lines $\beta_4 = \frac{27}{7}$. The discussion is similar to that for $\beta_2 = 3$.

(iii) $\beta_2 = 2\frac{1}{7}$. The equation for the surface $F_8 = 0$ simplifies. Here also the discussion must unfortunately be omitted owing to lack of space.

(iv) $\beta_2 = 3$. The section of the surface $F_2 = 0$ by the plane $\beta_2 = 3$ reduces to the two parallel straight lines $\beta_4 = 15$ and $7\beta_4 = 45$, i.e. $r = 0$ and $r = -10$, whereas the curves of section of the surface $F_6 = 0$ reduce to the straight line $\beta_4 = 15$ and the parabola $3\beta_6 - 275 = 49\beta_4^2 + 42\beta_4 + 9945$, i.e.

$$\Phi_6 = \frac{3\beta_6 - 275}{194} = \frac{r\,[600 + 36r + r^2]}{2716}, \text{ since } \beta_4 = 15 + \frac{6r}{7}.$$

Now

$$m_1 = 9\,\frac{(\beta_6 - \phi_4)}{(15 - \beta_4)},$$

$$q_1 p_1 = -\,\frac{9(\beta_6 - \phi_0)}{(15 - \beta_4)},$$

$$q_1 + p_1 = -\tfrac{1}{6}\,(7\beta_4 - 45),$$

$$(q_1 - p_1)^2 = \frac{1296\beta_6 - [49\beta_4^3 - 693\beta_4^2 + 9315\beta_4 - 30375]}{36\,(15 - \beta_4)},$$

$$m_1^2 - (q_1 - p_1)^2 = \frac{81}{4\,(15 - \beta_4)^2}\,(\beta_6 - \phi_8')(\beta_6 - \phi_8'').$$

The values of the ordinates Φ_0, Φ_4, Φ_6, Φ_8 are shown in Table XVI, p. 161, and the transition curves drawn in Diagrams XVI and XVII (right-hand figures).

We find that $F_6 = 0$ lies below both $F_0 = 0$ and $F_4 = 0$ for $\beta_4 < 15$ and above for $\beta_4 > 15$. Hence for points on $F_6 = 0$, $m_1 < 0$, $p_1 q_1 > 0$ and $q_1 + p_1 < 0$, and so we have the Type (vii) curve. Between the line $\beta_4 = 15$ and the curve $F_6 = 0$, $(q_1 - p_1)^2 < 0$ and we find a heterotypic area.

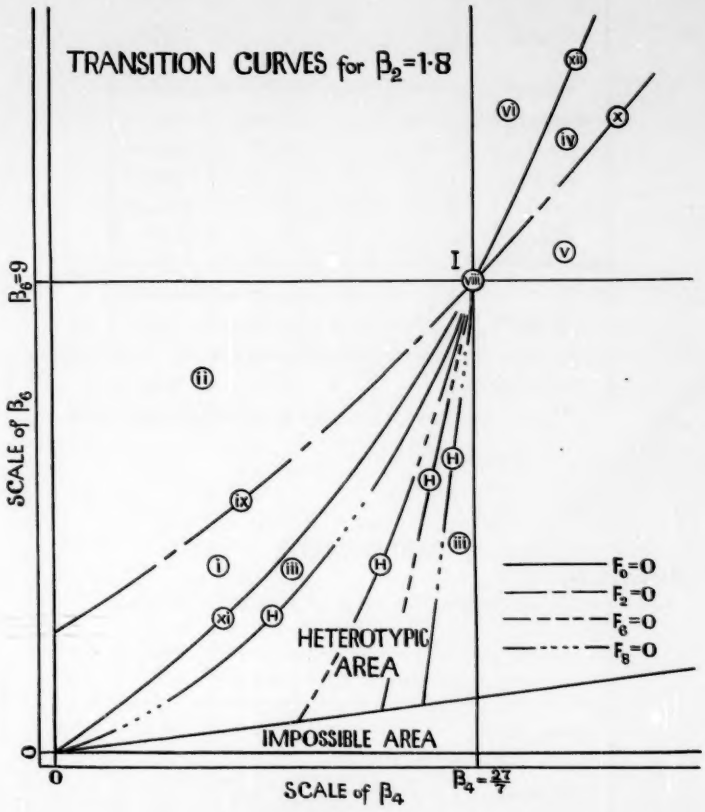


Diagram IV.

Again, the area which corresponds to the Type (iii) curve, viz. that between $F_0 = 0$ and $F_4 = 0$, always lies between the two branches of $F_8 = 0$ and hence the condition $\frac{m}{q_1 - p_1} < 1$ is always satisfied. Furthermore, points on $F_8' = 0$ correspond to the Type (xiii) curve when $\beta_4 < 15$ and to the Type (xv) when $\beta_4 > 15$; on the other hand, for $F_8'' = 0$ we first get the Type (xv) for $\beta_4 < 15$ and then the Type (xiii) for $\beta > 15$.

TABLE XXIII*a*.

(1) $\beta_4 < 10\frac{5}{7}$.

Zone	Signs of			Type
	m_1	$q_1 p_1$	$q_1 + p_1$	
Above $F_4=0$	+	−	−	(i)
On $F_4=0$	0	−	−	(viii)
Between $F_4=0$ and $F_0=0$	−	−	−	(iv)
On $F_0=0$	−	0	−	(xii)
Below $F_0=0$	−	+	−	(vi)

TABLE XXIII*b*.

(2) $10\frac{5}{7} < \beta_4 < 15$.

Zone	Signs of			Type
	m_1	$q_1 p_1$	$q_1 + p_1$	
Above $F_0=0$	+	−	−	(i)
On $F_0=0$	+	0	−	(xi)
Between $F_0=0$ and $F_4=0$	+	+	−	(iii)
On $F_4=0$	0	+	−	(viii)
Below $F_4=0$	−	+	−	(vi)

(3) $\beta_2 > 15$.

TABLE XXIII*c*.

Zone	Signs of			Type
	m_1	$q_1 p_1$	$q_1 + p_1$	
Above $F_4=0$	−	+	−	(vi)
On $F_4=0$	0	+	−	(viii)
Between $F_4=0$ and $F_0=0$	+	+	−	(iii)
On $F_0=0$	+	0	−	(xi)
Below $F_0=0$	+	−	−	(i)

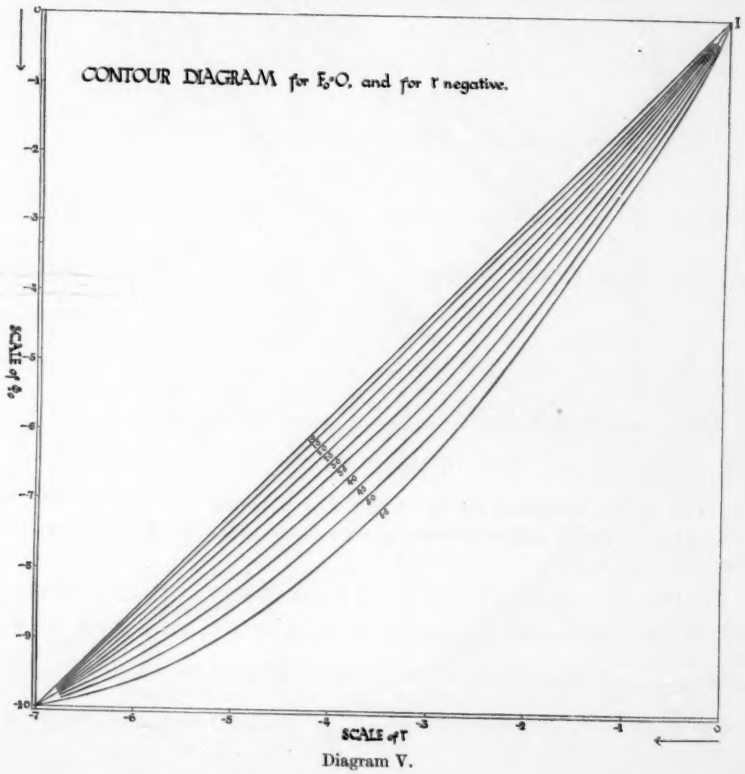


Diagram V.

(*g*) The contours of the transition surfaces.

In plotting the Φ-ordinate for different values of r, we find it convenient to make separate diagrams for $r > 0$ and $r < 0$.

(i) The contour diagrams for the surface $F_0 = 0$ are straightforward. Note that each curve of section must pass through the points $(-7, -1)$ and $(0, 0)$ and that Φ_0 decreases as β_2 increases. See Diagrams V and VI.

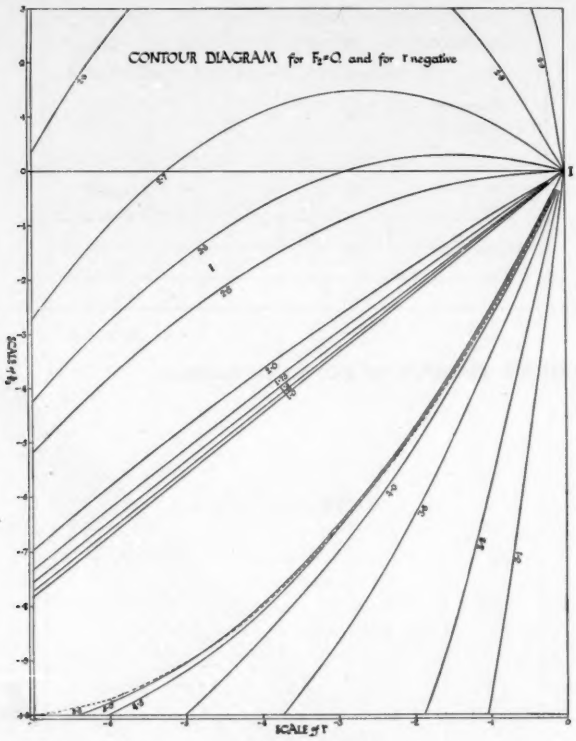


Diagram VI.

(ii) The contour diagrams for $F_2 = 0$ and $F_4 = 0$, shown in Diagrams VII, VIII and IX, present greater difficulty and we need more ordinates. As before we easily find the Φ-ordinates for

$r = -7 - (1{\cdot}0) - +7$ and $\beta_2 = 1{\cdot}7 - (\cdot 15) - 1{\cdot}9$ and $\beta_2 = 2{\cdot}6\,(-1) - 2{\cdot}9$, 3·1, and 3·2.

Let us now consider the contours of the surfaces of $F_2 = 0$ and $F_4 = 0$. When $r < 0$ ϕ_2 increases as β_2 increases till it becomes infinite at $\beta_2 = 3$. It now changes $r > 0$ decreases sign and as β_2 is further increased, ϕ_2 increases. On the other hand, for $r < 0$, ϕ_4 decreases

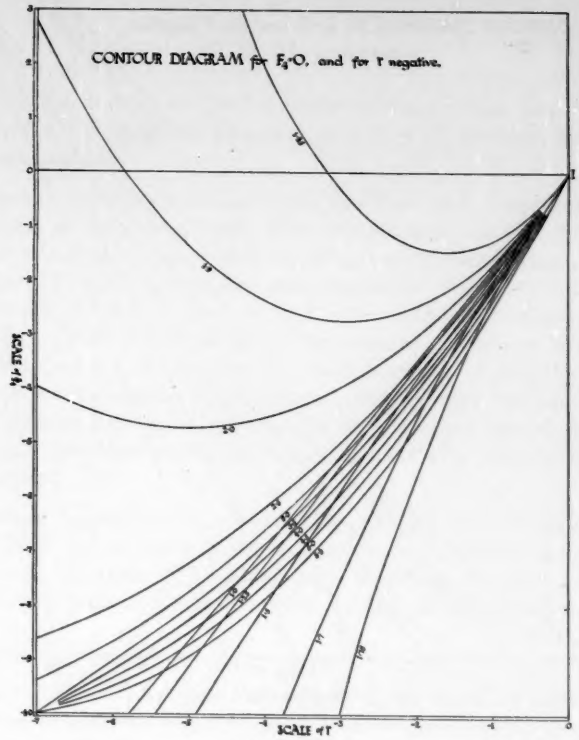


Diagram VII.

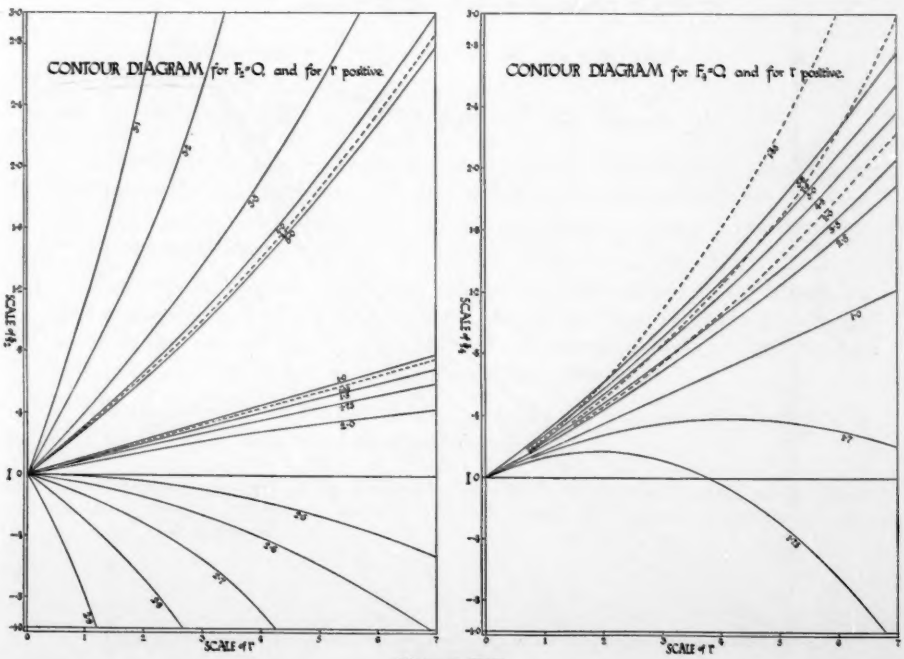


Diagram VIII.

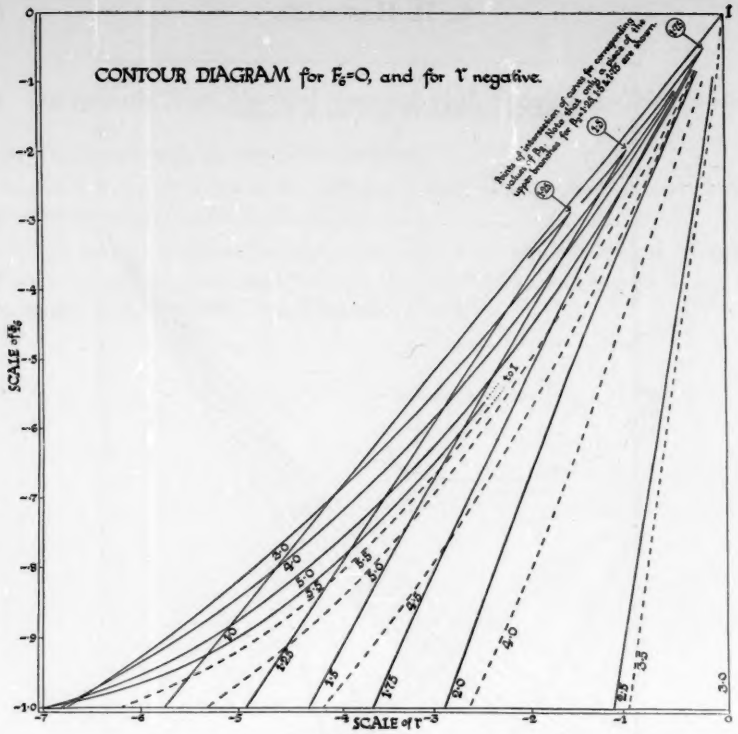

CONTOUR DIAGRAM for F_8=0, and for r negative.
Points of intersection of curves for corresponding values of P_8. Note that only pieces of the upper branches for P_8=1.25, 1.5, 1.75, 2.0 are shown.
SCALE of r^{-3}
SCALE of P_8

Diagram IX.

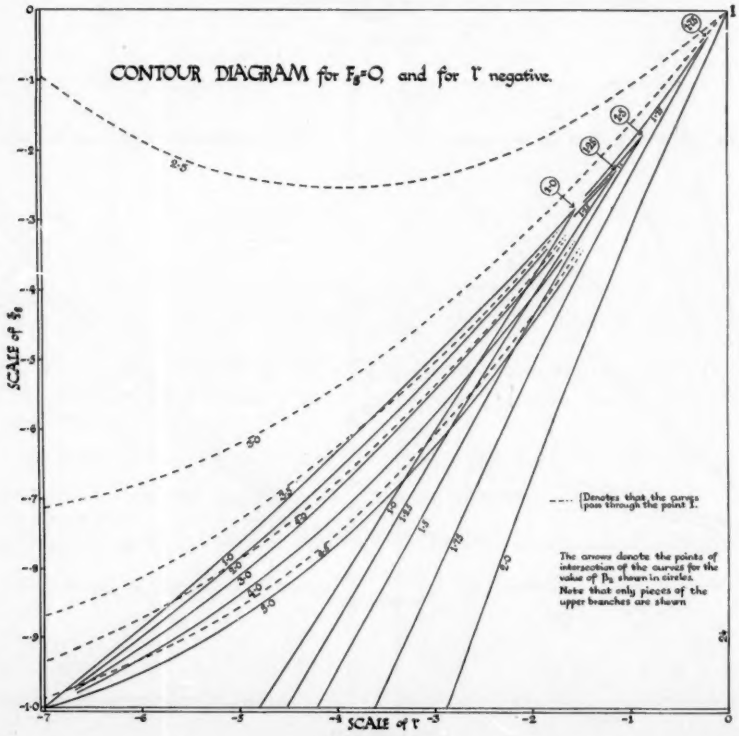

CONTOUR DIAGRAM for F_8=0, and for r negative.
Denotes that the curves pass through the point 1.
The arrows denote the points of intersection of the curves for the value of P_8 shown in circles.
Note that only pieces of the upper branches are shown
SCALE of r^{-3}
SCALE of P_8

Diagram X.

decreases, becomes infinite at $\beta_2 = 1{\cdot}8$, where it changes sign and then decreases again; but if $r > 0$, ϕ_4 decreases, changes sign at $\beta_2 = 1{\cdot}8$, decreases till $\beta_2 = 2{\cdot}5$ and then increases again.

The contours of the surfaces $F_6 = 0$ and $F_8 = 0$ are more complicated and I have omitted several of the lines to make their reproduction possible. Note that since the surfaces consist of two parts, each section gives two contour lines. If $r < 0$, the contour lines of $F_6' = 0$ are such that each member of the system cuts all the preceding members, e.g. the contour line for $\beta_2 = 2$ cuts the contour lines for $\beta_2 = 1{\cdot}75$, $\beta_1 = 1{\cdot}5$, and $\beta_2 = 1{\cdot}25$. Furthermore, ϕ_6'' decreases as β_2 increases, becomes infinite at $\beta_2 = 3$, but does not change sign and then increases. Again, the lower of the contour lines for the surface $F_8 = 0$ always passes through the points (0, 0) and $(-7, -1)$, whereas corresponding ordinates on the upper branch decrease until $\beta_2 = 2\frac{1}{7}$, where they become infinite, change sign and then decrease again, as β_2 is further increased.

The contour diagrams for $F_6 = 0$ and $F_8 = 0$ simplify for $r > 0$ because there are now fewer lines. ϕ_6' increases as β_2 increases, becomes infinite at $\beta_2 = 3$ and now decreases as β_2 increases. $F_6'' = 0$ represents a system of curves, such that each member cuts the preceding members. Both ϕ_6' and ϕ_6'' increase as β_2 increases.

(*h*) *The Transition Curves.*

Diagrams XIII to XXI show the sections of the transition surfaces by planes corresponding to different values of β_2 and the frequency types corresponding to the different areas*. These areas are determined as in Section I (*g*), p. 143, but I have now introduced the conditions

$$(q_1 - p_1)^2 = \frac{81\,(\beta_2 - 3)^2}{4K^2}(\beta_6 - \phi_6')(\beta_6 - \phi_6'') \geqq 0$$

and
$$m_1^2 - (q_1 - p_1)^2 = \frac{243\,(\beta_2 - 1)(7\beta_2 - 15)}{16K^2}(\beta_6 - \phi_8')(\beta_6 - \phi_8'') \leqq 0.$$

Accordingly the zones between the two branches of $F_6 = 0$ and also those parts of the zones for the main Types (iii), (iv), and (v) and the transition Types (x) where $m_1^2 - (q_1 - p_1)^2$ is not negative, are now heterotypic areas, points of which do not satisfy the condition $c_a = 0$.

Furthermore, note that the Type (vii) can only occur when $F_6 = 0$ borders an area corresponding to the Type (vii) curve. Again, the Types (xiii), (xiv), (xv) must occur in the areas of the Type (i), (ii), and (vi) curves respectively. At all other points of $F_6 = 0$ and $F_8 = 0$ we get a different type of curve.

Clearly $m_1^2 - (q_1 - p_1)^2 < 0$ if we are outside the two branches of F_8 when $\beta_2 < 1\frac{5}{7}$ or if we are inside these two branches when $\beta_2 > 1\frac{5}{7}$.

* The heterotypic areas are the zones where we may get other types of curves, but not those considered above, and the impossible areas are those wherein the necessary inequalities $\beta_4 > \beta_2^3$, $\beta_6 > \beta_2\beta_4 > \beta_2^3$ do not hold.

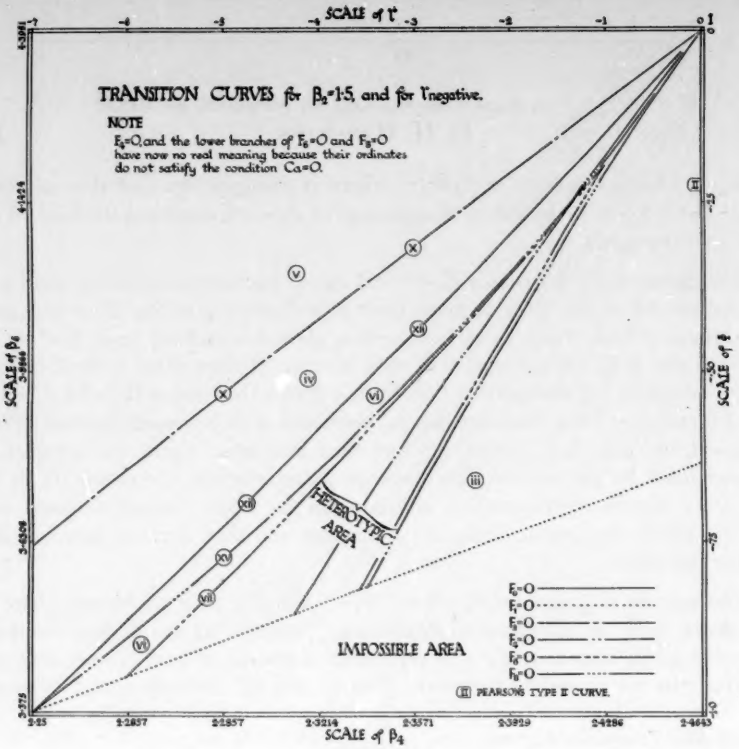


Diagram XI.

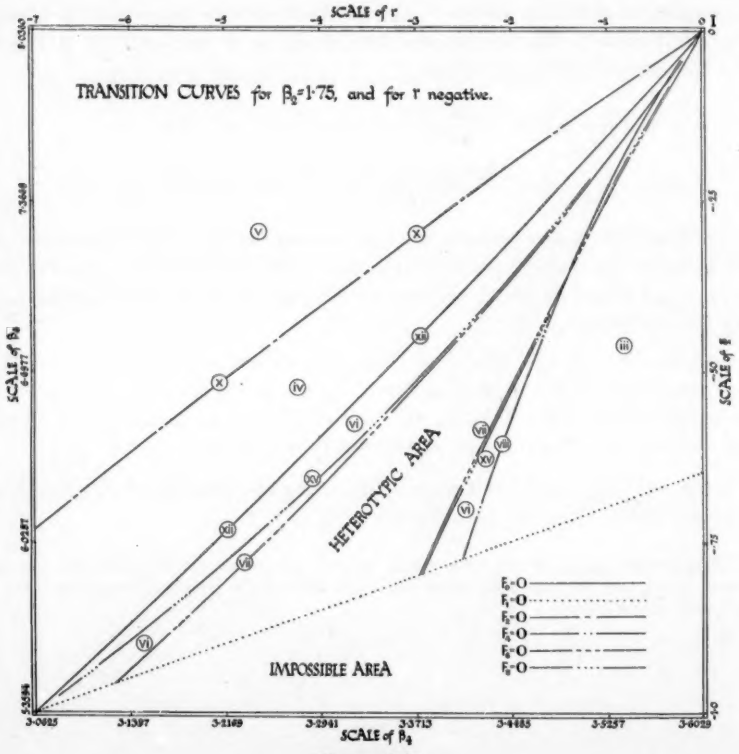


Diagram XII.

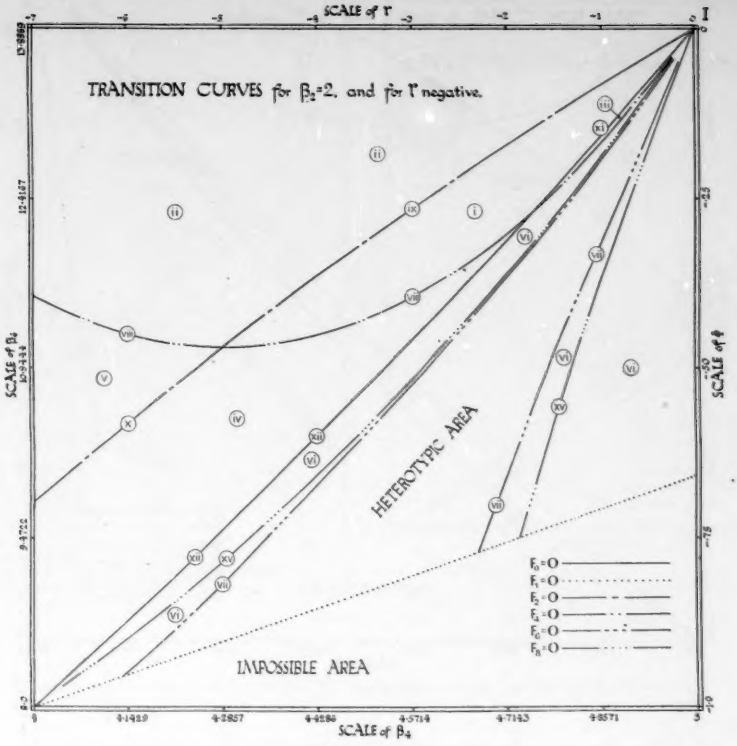


Diagram XIII.

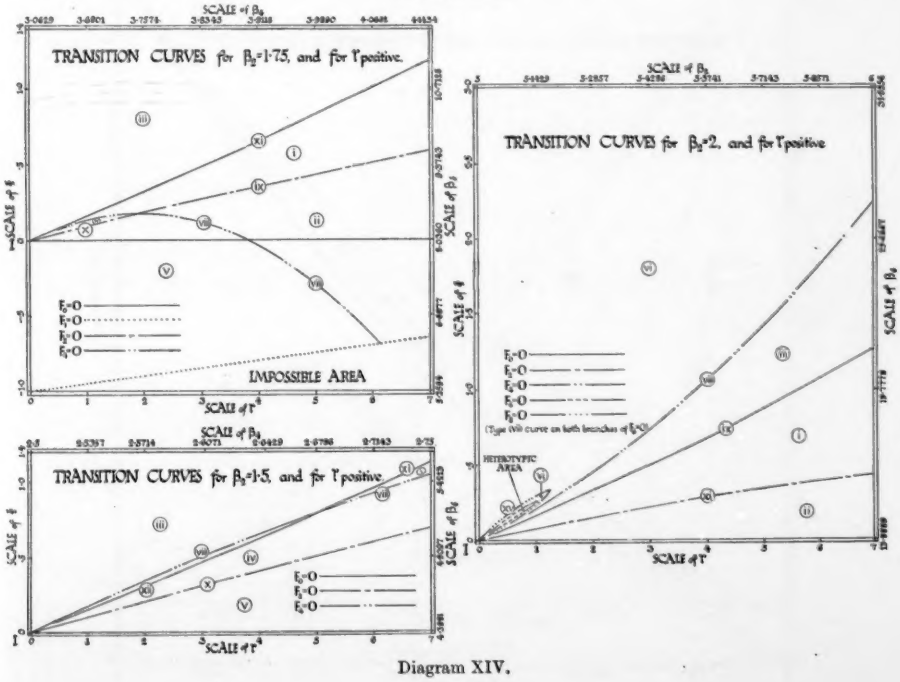

TRANSITION CURVES for β_3=1·75, and for I' positive.
IMPOSSIBLE AREA
TRANSITION CURVES for β_3=1·5, and for I' positive.
TRANSITION CURVE for β_3=2, and for I' positive
HETEROTYPIC AREA
SCALE of I'
SCALE of β_3
SCALE of β_4

Diagram XIV.

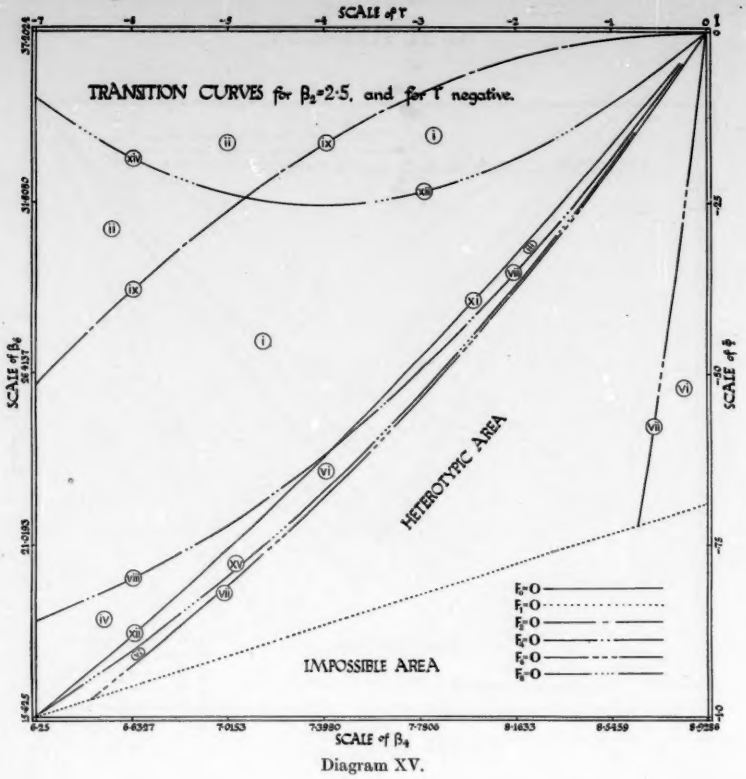


Diagram XV.

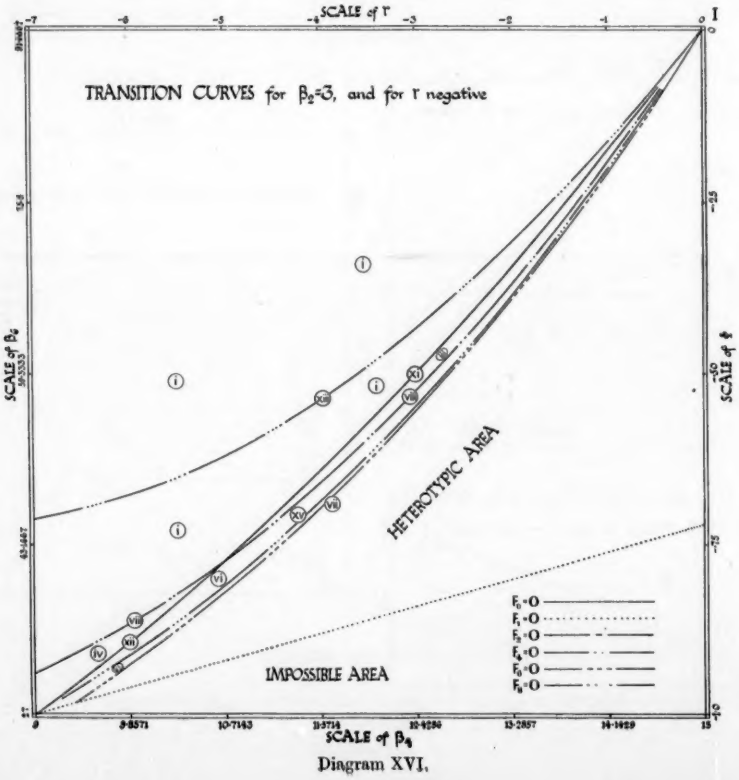


Diagram XVI.

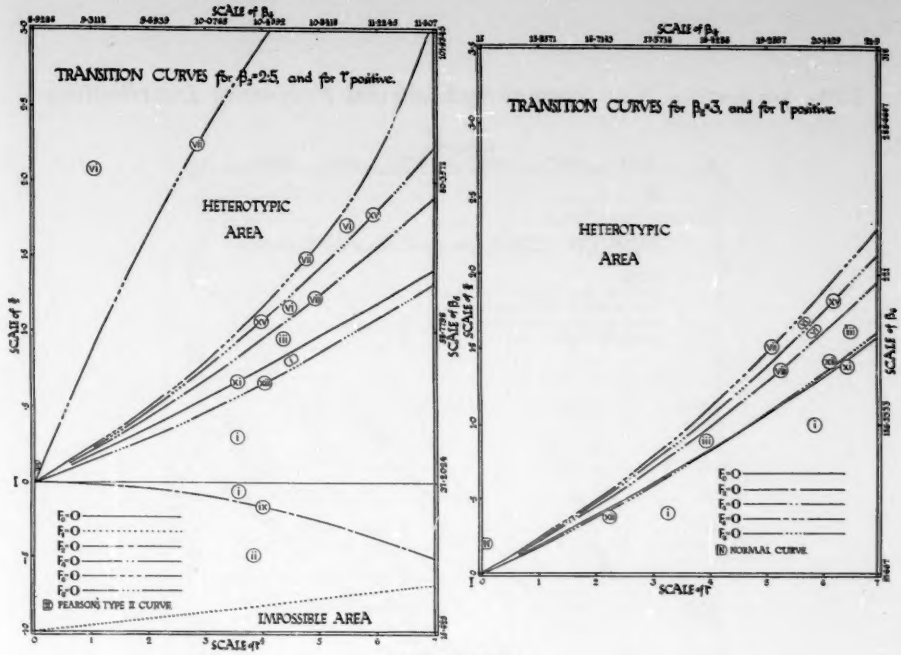


Diagram XVII.

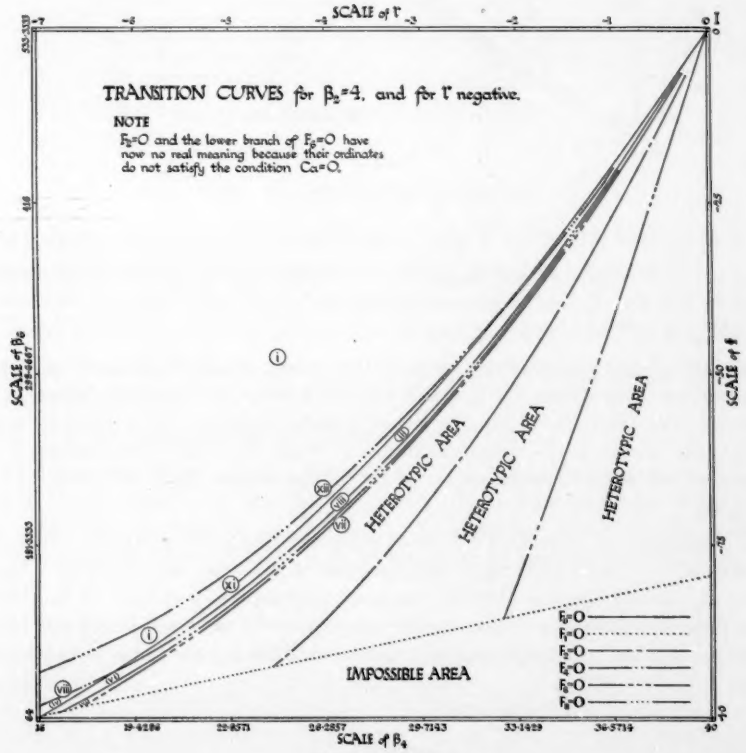


Diagram XVIII.

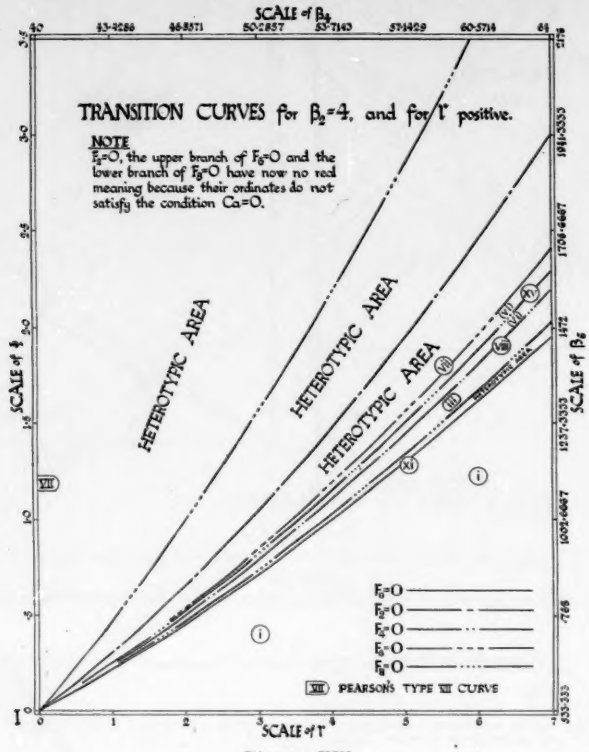


Diagram XIX.

Note (1) that the sign of K does not trouble us.

(2) When $F_6'=0$ and $F_6''=0$ become imaginary, $(q_1-p_1)^2$ is always positive; when $F_8'=0$ and $F_8''=0$ become imaginary, $m_1{}^2-(q_1-p_1)^2$ is positive or negative according as $\beta_2 >$ or $< \frac{1.5}{7}$.

In (f) of the first Section (see p. 141) I gave the values of β_4 and β_6 at the intersections of the transition surfaces with each other, but all these values do not come into the diagrams of the transition surfaces because I have given an upper limit to β_4. Table XXIV gives the values of r and Φ at the intersections of the transition curves for different values of β_2. These values check the Diagrams XI to XIX.

By painting the transition curves on glass and fixing the sheets corresponding to each value of β_2 in their right order, one above the other, we can make a model in the β_2, β_4 and β_6-space showing the zones corresponding to each of the fourth order Pearson curves. But unfortunately the curves are so close to each other that the model becomes too complicated and its reproduction in print would be impossible.

TABLE XXIV.

β_2	Surfaces which intersect	Co-ordinates of points of intersection	
		r	Φ
1·5	$F_8''=0$, $F_6'=0$, and $F_4=0$ $F_0=0$ and $F_4=0$	−1·302 5·0	−·2437 +·7919
1·75	$F_8''=0$, $F_6'=0$ and $F_4=0$ $F_6''=0$ and $F_4=0$ $F_0=0$ and $F_4=0$ $F_2=0$ and $F_4=0$	−1·333 −·409 +·556 +1·905	−·3326 −·0835 +·0877 +·1759
2·0	$F_0=0$ and $F_4=0$ $F_2=0$ and $F_4=0$	−1·667 −5·0	−·2621 −·4716
2·5	$F_8''=0$ and $F_2=0$ $F_0=0$ and $F_4=0$	−4·807 −3·889	−·2439 −·6066
3·0	$F_8''=0$ and $F_0=0$ $F_0=0$ and $F_4=0$	+5·0 −5·0	+1·0677 −·7732
4·0	$F_8'=0$ and $F_0=0$ $F_0=0$ and $F_4=0$	−1·667 −6·111	−·3247 −·9259

III. STATISTICAL APPLICATIONS.

(*a*) *General Remarks on Curve Fitting.*

A model would enable us to determine by inspection what type of curve must be used given the values of β_2, β_4 and β_6. This method is not practicable, because the model cannot readily be produced with sufficient accuracy for this purpose.

In the illustrations dealt with in this Section, I have shown how to determine the curve to be fitted:

(1) from first principles as in Example (*d*) on p. 188;

(2) by interpolating into the Tables XI—XXI and using Diagram I, Figs. I, II and III, to determine the type of curve, then finding if the point falls into a heterotypic area or corresponds to a transition curve on the surfaces $F_6=0$ or $F_8=0$ by considering the contour lines of these surfaces as in Example (*e*), p. 191;

(3) by inspection of the two neighbouring transition curve diagrams, etc., as in Example (*f*), p. 193.

I have discussed the transition curves at great length for $\beta_2 < 6$. For higher values of β_2 we must calculate the type of curve to be used from first principles, following the method used in Example (*d*), p. 188.

(*b*) *Distribution of Product Moment Coefficients, p_{11}.*

The distribution of the first product moment coefficient in samples drawn from an indefinitely large normal population is a Bessel-function* curve. When the sample is larger than 25 an excellent fit is obtained by using a Pearson curve with the same mean, standard deviation, and constants β_1 and β_2. For smaller samples a Pearson curve cannot be used to replace the Bessel-function curve. The distribution of the first product moment coefficient is symmetrical when the variates of the population sampled are uncorrelated. The problem arises whether one of my curves with the same mean, the same standard deviation and the same constants β_2, β_4 and β_6 can describe the corresponding symmetrical Bessel-function curve.

The distribution of $v = p_{11} \dfrac{n}{1-\rho^2} \dfrac{1}{\sigma_1 \sigma_2}$ is

$$y = \frac{N(1-\rho^2)^{\frac{n-1}{2}}}{\sqrt{\pi} \cdot 2^{\frac{n-2}{2}} \Gamma\left(\frac{n-1}{2}\right)} e^{\rho v} \left\{ v^{\frac{n-2}{2}} K_{\frac{n-2}{2}}(v) \right\}$$

$$= \frac{N(1-\rho^2)^{\frac{n-1}{2}}}{\sqrt{\pi} \cdot 2^{\frac{n-2}{2}} \Gamma\left(\frac{n-1}{2}\right)} e^{\rho v} T_m(v), \qquad m = \frac{n-2}{2},$$

where n is the size of the sample, σ_1, σ_2 and ρ refer to the parent population. Log $T_m(v)$ is tabled* for $m = 0$ to $m = 11{\cdot}5$ and for different values of v.

The sth moment coefficient* of p_{11} about zero is

$$\mu_s' = \frac{I_s}{I_0},$$

where

$$I_s = A_s \int_0^\infty [e^{2\rho u} + (-1)^s e^{-2\rho u}] u^s F(u)\, du,$$

and

$$F(u) = \frac{1}{2} \Gamma\left(\frac{n-2}{2}\right) u^{\frac{n-2}{2}} K_{\frac{n-2}{2}}(2u).$$

A_s is found to be

$$A_s = \frac{N 2^{n-2} (1-\rho^2)^{\frac{n-1}{2}}}{\pi \Gamma(n-2)} [2\sigma_1 \sigma_2 (1-\rho^2)]^s.$$

On evaluating I_s and dividing by I_0 we find

$$\mu_s' = 2 \left(\frac{\sigma_1 \sigma_2}{n}\right)^s \frac{\Gamma(n+s-1)\, \Gamma(s+1)}{\Gamma\left(\frac{n+2s+1}{2}\right) \Gamma\left(\frac{n-1}{2}\right)}.$$

* *Biometrika*, Vol. xxi. p. 164, *sq.*

The reduction formula for $f_s(\rho)$ is

$$f_s(\rho) = \frac{1}{2}\left(\frac{n-1}{2} + s\right)\int f_{s-1}(\rho)\, d\rho + \text{constant}.$$

The constant of integration is found from the two reductions

$$f_s(0) = 0 \text{ for odd } s,$$

and

$$f_s(1) = \frac{\Gamma\left(\frac{n+1}{2}+s\right)\Gamma\left(\frac{n-1}{2}+s\right)}{\Gamma(n+s-1)\,\Gamma(s+1)}.$$

$f_{-1}(\rho) = 0$, and hence we can find $f_r(\rho)$ $(r = 0, 1, \ldots 8)$*.

We find

$$f_5(\rho) = \frac{1}{768}\frac{\Gamma\left(\frac{n+11}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}{\Gamma(n-1)}\left[\frac{\rho^5}{5} + \frac{2\rho^3}{n} + \frac{3\rho}{n(n+2)}\right],$$

$$f_6(\rho) = \frac{1}{46080}\frac{\Gamma\left(\frac{n+13}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}{\Gamma(n-1)}\left[\rho^6 + \frac{15\rho^4}{n} + \frac{45\rho^2}{n(n+2)} + \frac{15}{n(n+2)(n+4)}\right],$$

$$f_7(\rho) = \frac{1}{92160}\frac{\Gamma\left(\frac{n+15}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}{\Gamma(n-1)}\left[\frac{\rho^7}{7} + \frac{3\rho^5}{n} + \frac{15\rho^3}{n(n+2)} + \frac{15\rho}{n(n+2)(n+4)}\right],$$

$$f_8(\rho) = \frac{1}{10321920}\frac{\Gamma\left(\frac{n+17}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}{\Gamma(n-1)}\left[\rho^8 + \frac{28\rho^6}{n} + \frac{210\rho^4}{n(n+2)} + \frac{420\rho^2}{n(n+2)(n+4)} + \frac{105}{n(n+2)(n+4)(n+6)}\right].$$

Put $\rho = 0$, then the sth moment about the mean is

$$\mu_s = \left(\frac{2\sigma_1\sigma_2}{n}\right)^s \frac{\Gamma(n+s-1)\,\Gamma(s+1)}{\Gamma\left(\frac{n+2s+1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} f_s(0).$$

Putting $s = 2$, 4, 6, 8 and substituting for $f_s(0)$, we can determine μ_s, and therefore the corresponding β's. We easily find

$$\beta_2 = \frac{3(n+1)}{n-1},$$

$$\beta_4 = \frac{15(n+1)(n+3)}{(n-1)^2},$$

$$\beta_6 = \frac{105(n+1)(n+3)(n+5)}{(n-1)^3}.$$

* $f_0(\rho)$, $f_1(\rho)$, $f_2(\rho)$, $f_3(\rho)$ and $f_4(\rho)$ are worked out *loc. cit.*—the higher functions follow after some lengthy algebra.

Hence from the equations of (*b*), p. 136, we find

$$p_1 q_1 = -\frac{(n+1)(n+3)(n+15)}{n-1} = -qp,$$

$$p_1 + q_1 = -\frac{n^2+28n+90}{n-1} = -(q-p),$$

$$m_1 = \frac{n^3+22n^2+132n+300}{2(n-1)} = m,$$

where $q=|q_1|$, $p=|p_1|$, $m=|m_1|$, $q>p$. Therefore $q=-q_1$, $p=p_1$, $m=m_1$.

Similarly

$$k' = \frac{m}{q+p} = \frac{1}{2}\frac{n^3+22n^2+132n+300}{\sqrt{(5n^4+128n^3+1140n^2+4968n+7920)}}$$

on simplification. The size of the sample, n, is >0, and therefore my Type (i) curve must always be used. The equation is

$$y = y_0\left(\frac{1-\frac{u^2}{q}}{1+\frac{u^2}{p}}\right)^{k'},$$

where

$$u = \frac{v}{\sigma_v} = \frac{v}{\sqrt{(n-1)}}.$$

Taking the total frequency to be 1000 in each case, I have calculated the ordinates and the frequencies for samples of size $n=3$, 5, 10, 25 and 50, for the Type VII curve, the normal curve, and my Type (i) curve. I compared the values found with the corresponding values of $y=1000\,T_m$, as found from the tables, the ordinates by plotting, and the frequencies by using the χ^2-test for goodness of fit. Strictly speaking, we cannot use the χ^2-test in this case, because the test assumes an observed distribution, found in sampling. I used it because I know of no better test for comparing the frequencies.

For the Type (i) curve, I found y_0 by quadrature. I took

$$y_0'' = c_{\frac{n-2}{2}}(0) = 1000\,T_m(0)$$

to find the ordinates, and then found the correcting factor (y_0') which would reduce the area under the curve to 1000. Thus the true value for

$$y_0 = (y_0') \times (y_0'') = (y_0') \times 1000\,T_{\frac{n-2}{2}}(0).$$

The equations at the top of this page give the following values for the constants of the curves:

TABLE XXV.

n	p	q	k'	y_0'	$c_{\frac{n-2}{2}}$
3	2·3027	93·8027	2·3958	0·6995	500·00
5	3·5653	67·3153	2·8834	0·9495	250·00
10	6·7372	58·9594	4·0760	0·9864	145·51
25	16·1536	75·1120	7·5273	0·9985	84·09
50	31·6961	113·1247	13·1690	1·0004	57·83

Professor Pearson's Type VII curve, viz.

$$y = y_0 \frac{1}{\left(1+\frac{v^2}{a^2}\right)^l},$$

where $$y_0 = \frac{N}{\sqrt{\{\pi\,(n^2-1)\}}} \frac{\Gamma\left(\frac{n+4}{2}\right)}{\Gamma\left(\frac{n+3}{2}\right)}, \quad a^2 = n^2-1, \quad l = \tfrac{1}{2}\,(n+4),$$

gives the following constants:

TABLE XXVI.

n	a^2	l	y_0
3	8	3·5	331·45
5	24	4·5	223·26
10	99	7·0	141·81
25	624	14·5	83·76
50	2499	27·0	57·83

The normal curve is $$y = \frac{1000}{\sigma\sqrt{(2\pi)}}\, e^{-\frac{1}{2}\frac{v^2}{n-1}}.$$

I calculated the ordinates and the frequencies for each of the three types when $n = 3, 5, 10, 25, 50$. The ordinates when $n = 5$ and $n = 10$ are plotted in Diagrams XX and XXI. The values of P are approximately equal for the Type (i) and the Type VII curves, but the normal curve fits badly. The values of χ^2 however show that the Type (i) can better describe the Bessel-function curve than the Type VII, and this is also shown by the diagrams. For samples of 10 the Type (i) curve effectively replaces the Bessel-function curve, for values of $n > 25$ the Type VII and the Type (i) curves become identical.

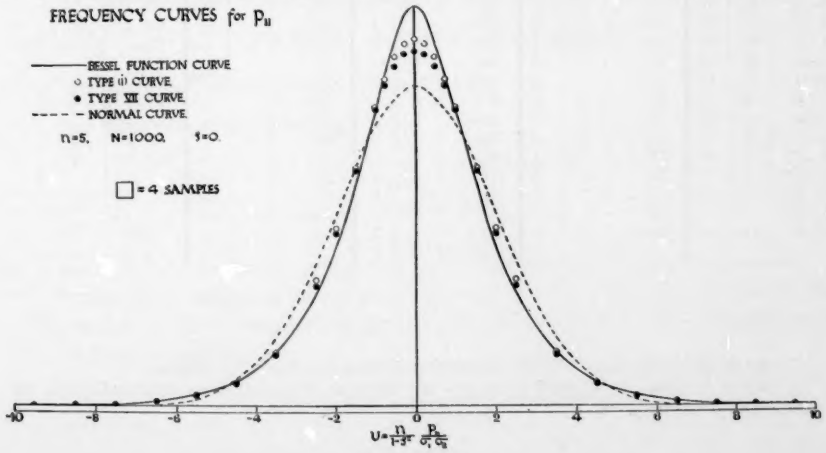


Diagram XX.

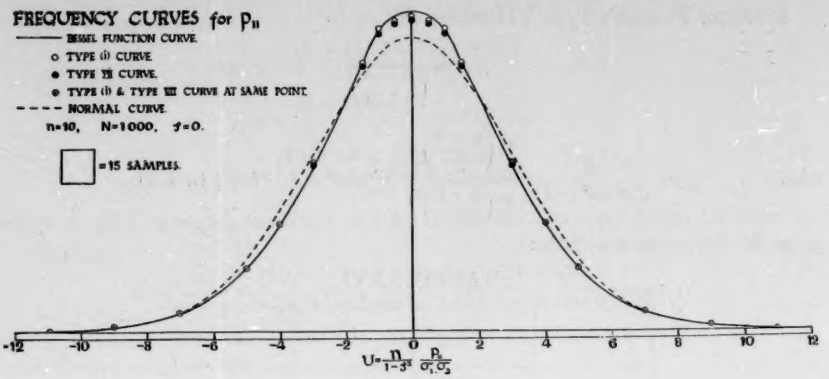


Diagram XXI.

(c) *Distribution of μ_3 in samples.*

Dr J. Pepper* drew samples of 10 from a normal paper with variance 100 and calculated the 3rd moment coefficient of each sample. He found the following distribution:

TABLE XXVII†.

Group	Frequency +	Frequency −	Group	Frequency +	Frequency −	Group	Frequency +	Frequency −	Group	Frequency +	Frequency −	Group	Frequency +	Frequency −
0—1	42	49	15—	7	4	30—	—	—	45—	—	1	60—	—	—
1—2	42	31	16—	7	15	31—	—	—	46—	—	—	61—	—	—
2—3	42	36	17—	5	3	32—	1	1	47—	—	—	62—	—	—
3—4	27	21	18—	4	2	33—	1	1	48—	—	—	63—	—	—
4—5	26	28	19—	3	2	34—	2	1	49—	—	—	64—	—	—
5—6	28	17	20—	4	3	35—	2	—	50—	—	—	65—	—	—
6—7	15	17	21—	5	2	36—	—	—	51—	—	1	66—	—	—
7—8	19	12	22—	4	2	37—	2	—	52—	—	—	67—	—	—
8—9	15	15	23—	1	7	38—	1	—	53—	—	—	68—	—	—
9—10	11	2	24—	2	4	39—	—	—	54—	—	—	69—	—	—
10—11	9	16	25—	1	1	40—	—	—	55—	—	—	70—71	1	1
11—12	7	8	26—	1	4	41—	—	—	56—	—	—			
12—13	4	3	27—	1	4	42—	—	3	57—	—	—			
13—14	7	8	28—	1	—	43—	1	—	58—	—	—			
14—15	11	8	29—	3	1	44—	1	—	59—	—	—			

Total	Frequency +	Frequency −
	366	334
	700	

* I am greatly indebted to Dr Pepper for placing his sampling data at my disposal.

† I denote by Frequency + and Frequency − the frequency in groups to the right and to the left of the origin respectively.

We want Sheppard's correction formulae for the higher μ's. Now the general formula* is

$$\nu_p = \mu_p + \frac{h^2}{3!\,2^2} p(p-1)\mu_{p-2} + \frac{h^4}{5!\,2^4} p(p-1)(p-2)(p-3)\mu_{p-4} + \ldots,$$

where $\nu_p = p$th raw moment coefficient about the mean and $\mu_p =$ corrected pth moment coefficient.

Hence

$$\left.\begin{aligned}
\mu_5 &= \nu_5 - h^2 \cdot \tfrac{5}{6}\mu_3 \\
\mu_6 &= \nu_6 - h^2 \cdot \tfrac{5}{4}\mu_4 - h^4 \cdot \tfrac{3}{16}\mu_2 - \tfrac{1}{448}h^6 \\
\mu_7 &= \nu_7 - h^2 \cdot \tfrac{7}{4}\mu_5 - h^4 \cdot \tfrac{7}{16}\mu_3 \\
\mu_8 &= \nu_8 - h^2 \cdot \tfrac{7}{3}\mu_6 - h^4 \cdot \tfrac{7}{8}\mu_4 - h^6 \cdot \tfrac{1}{16}\mu_2 - \frac{h^8}{2304}
\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots(15).$$

Again, if ν_p' denotes the uncorrected moment coefficient about the origin, then

$$\nu_p = \nu_p' - p\nu'_{p-1}\nu_1' + \frac{p(p-1)}{2!}\nu'_{p-2}\nu_1'^2 - \ldots.$$

Hence

$$\nu_5 = \nu_5' - 5\nu_4'\nu_1' + 10\nu_3'\nu_1'^2 - 10\nu_2'\nu_1'^3 + 4\nu_1'^5,$$

$$\nu_6 = \nu_6' - 6\nu_5'\nu_1' + 15\nu_4'\nu_1'^2 - 20\nu_3'\nu_1'^3 + 15\nu_2'\nu_1'^4 - 5\nu_1'^6,$$

$$\nu_7 = \nu_7' - 7\nu_6'\nu_1' + 21\nu_5'\nu_1'^2 - 35\nu_4'\nu_1'^3 + 35\nu_3'\nu_1'^4 - 21\nu_2'\nu_1'^5 + 6\nu_1'^7,$$

$$\nu_8 = \nu_8' - 8\nu_7'\nu_1' + 28\nu_6'\nu_1'^2 - 56\nu_5'\nu_1'^3 + 70\nu_4'\nu_1'^4 - 56\nu_3'\nu_1'^5 + 28\nu_2'\nu_1'^6 - 7\nu_1'^8.$$

The following constants were found:

$$\mu_1 = \cdot1957, \quad \sigma = 12\cdot334, \quad \beta_1 = \cdot0102,$$

$$\beta_2 = 7\cdot538, \quad \beta_4 = 139\cdot04, \quad \beta_6 = 3702\cdot24,$$

$$p_1 + q_1 = -57\cdot815 = -(q-p), \quad p_1 q_1 = -40\cdot208 = -qp, \quad m_1 = 87\cdot980 = m;$$

therefore $$p = \cdot6873, \quad q = 58\cdot50, \quad k' = 1\cdot4864.$$

The range is from $x = -94\cdot3$ to $x = +94\cdot3$.

The Type (i) curve accordingly is

$$y = y_0' \left\{ \frac{1 - \dfrac{x^2}{8899\cdot45}}{1 + \dfrac{x^2}{104\cdot55}} \right\}^{1\cdot4864}.$$

I calculated ordinates at $x = 0 - (\cdot125) - \cdot5$, $x = 1\cdot5 - (2) - 95$ and found the frequencies in the corresponding groups. The fit is very bad. The frequencies in the first twelve groups are:

* Whittaker and Robinson, *Calculus of Observations*, p. 196.

Group	Theoretical	Observed +	Observed −
0—1	35·1	42	49
1—2	34·2	42	31
2—3	32·3	42	36
3—4	29·8	27	21
4—5	27·0	26	28
5—6	24·1	28	17
6—7	21·2	15	17
7—8	18·5	19	12
8—9	16·0	15	15
9—10	13·8	11	2
10—11	11·9	9	16
11—12	10·3	7	8

Pepper also obtained the values for higher β's of the distribution of the 3rd moment in samples*, viz.

$$\left.\begin{aligned}
\beta_2 &= \frac{3(n^2+27n-70)}{(n-1)(n-2)} \\
\beta_4 &= \frac{15(n^4+84n^3+2695n^2-15{,}168n+20{,}020)}{(n-1)^2(n-2)^2} \\
\beta_6 &= \frac{105\{n^6+171n^5+13{,}893n^4+580{,}401n^3-5{,}131{,}014n^2 \\ \qquad +14{,}132{,}268n-12{,}932{,}920\}}{(n-1)^3(n-2)^3}
\end{aligned}\right\} \quad \ldots(16).$$

Hence substituting for β_2, β_4, β_6 in the equations giving p_1q_1, p_1+q_1 and m_1 we obtain, after some very lengthy algebra,

$$p_1q_1 = \frac{14\beta_4^2-9\beta_2\beta_6-5\beta_2^2\beta_4}{5\beta_2^2-\beta_4(6-\beta_2)}$$

$$= -\frac{1}{18}\frac{[A]}{[B]} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(17\,a),$$

where $$A = n^8+348n^7+50{,}975n^6+4{,}064{,}292n^5+34{,}629{,}558n^4$$
$$-646{,}758{,}090n^3+2{,}928{,}581{,}096n^2$$
$$-5{,}507{,}413{,}800n+3{,}806{,}202{,}400,$$

and $$B = (n-1)(n-2)(3n^4+642n^3-4879n^2+12{,}228n-10{,}220),$$

$$p_1+q_1 = \frac{9\beta_6(3-\beta_2)+7\beta_4^2-50\beta_2\beta_4+25\beta_2^3}{2\{5\beta_2^2-\beta_4.6-\beta_2\}}$$

$$= -\frac{1}{12}\frac{[C]}{[B]} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(17\,b),$$

* *Biometrika*, Vol. XXIV. p. 60.

where
$$C = 10n^7 + 3546n^6 + 521{,}736n^5 + 34{,}738{,}968n^4 - 392{,}792{,}064n^3 + 1{,}575{,}584{,}616n^2 - 2{,}789{,}485{,}552n + 1{,}854{,}410{,}880.$$

Further
$$m_1 = \frac{9\beta_6(5\beta_2 - 9) - 49\beta_4^2 + 210\beta_2\beta_4 - 125\beta_2^3}{4(6\beta_2^2 - 6 - \beta_2\beta_4)} = \frac{1}{36}\frac{[D]}{[B]} \quad \text{.....(17 c),}$$

where
$$D = n^8 + 393n^7 + 66{,}122n^6 + 6{,}216{,}894n^5 + 187{,}648{,}764n^4 - 2{,}369{,}922{,}228n^3 + 9{,}840{,}116{,}048n^2 - 17{,}754{,}485{,}784n + 11{,}957{,}893{,}360.$$

For all positive integral values of n the signs of p_1q_1, $q_1 + p_1$ and m_1 are negative, negative, and positive respectively. Hence we find in all cases that the Type (i) curve has the same first eight moment coefficients as the actual curve of distribution of μ_3.

Let us regroup the data, in groups of 5 times Pepper's grouping unit, and compare the distribution so obtained with the curves fitted from the theoretical values and observed values of β_2, β_4 and β_6.

The regrouped distribution is

TABLE XXVIII.

Group	Frequency		Group	Frequency		Group	Frequency		Group	Frequency		Group	Frequency	
	+	−		+	−		+	−		+	−		+	−
0—5	179	165	15—20	26	26	30—35	4	3	45—50	—	1	60—65	—	—
5—10	88	63	20—25	16	18	35—40	5	—	50—55	—	1	65—70	—	—
10—15	38	43	25—30	7	10	40—45	2	3	55—60	—	—	70—75	1	1

We want to have the standard deviation of the curve calculated from the theoretical betas expressed in the same units as the standard deviation of the distribution above.

Now
$$\sigma^2_{\mu_3} = \frac{6(n-1)(n-2)}{n^3}\mu_2^3,$$

and μ_2 was taken = 100 and the size of the sample is $n = 10$. Hence

$$\sigma^2_{\mu_3} = \frac{6 \cdot 9 \cdot 8}{10^3} \cdot 10^6 = 432{,}000$$

in actual units. Now the interval between successive groups in the distribution is $5\sigma = 50$ and therefore

$$\sigma^2_{\mu_3} = \frac{432{,}000}{2500} = 172{\cdot}8.$$

Put $n = 10$ in the Equations (17 *a*), (17 *b*) and (17 *c*) and we obtain

$$q_1 + p_1 = -554{\cdot}69, \quad \text{therefore } p = 1{\cdot}8823,$$
$$q_1 p_1 = -1047{\cdot}65, \quad \text{,, } q = 556{\cdot}572,$$
$$m_1 = 2076{\cdot}89, \quad \text{,, } k' = 2{\cdot}372.$$

Hence the curve required is

$$y = 28{\cdot}183 \left(\frac{1 - x^2 \times {\cdot}00001040}{1 + x^2 \times {\cdot}003074}\right)^{2{\cdot}372}.$$

The regrouped data corresponds to the Type (i) curve, viz.

$$y = 182{\cdot}37 \left(\frac{1 - x^2 \times {\cdot}00010442}{1 + x^2 \times {\cdot}010480}\right)^{1{\cdot}4708}.$$

The following frequencies were found:

TABLE XXIX.

Group	Theoretical*		Observed		Group	Theoretical		Observed	
	(*a*)	(*b*)	+	−		(*a*)	(*b*)	+	−
0— 5	132·9	162·5	179	165	50—55	·64	·70	—	1
5—10	96·4	92·9	88	63	55—60	·43	·51	—	—
10—15	56·0	43·9	38	43	60—65	·29	·35	—	—
15—20	29·5	21·5	26	26	65—70	·20	·23	—	—
20—25	14·7	11·4	16	18	70—75	·15	·15	1	1
25—30	8·1	6·6	7	10	75—80	·11	·09	—	—
30—35	4·5	4·0	4	3	80—85	·08	·05	—	—
35—40	2·6	2·5	5	—	85—90	·06	·03	—	—
40—45	1·6	1·7	2	3	χ^2	46·28	20·68	$n = 14$	
45—50	1·0	1·1	—	1	P	·00006	·08		

* The brackets show the grouping for the χ^2 test. (*a*) and (*b*) refer respectively to the theoretical curve and the curve from the data.

The poor fit when we fit the theoretical curve to the observed data is due to the large difference in the β's which seems to suggest that the total number of samples, viz. 700, is far too small, e.g.

β_2	β_4	β_6	Calculated from
7·538	139·04	3702	Data in Table XXVII
7·640	150·23	4257	Data in Table XXVIII
12·5	670·8	99225	Formulae (17 *a*—*c*)

Although $P = {\cdot}08$ for the curve fitted from the data, this is not a bad fit, because the contribution to χ^2 of the frequency in the group -5 to -10, viz. 63 against a theoretical value of 92·9, is 9·6. Diagram XXII gives the regrouped distribution with the ordinates of the corresponding Type (i) and Type II curves.

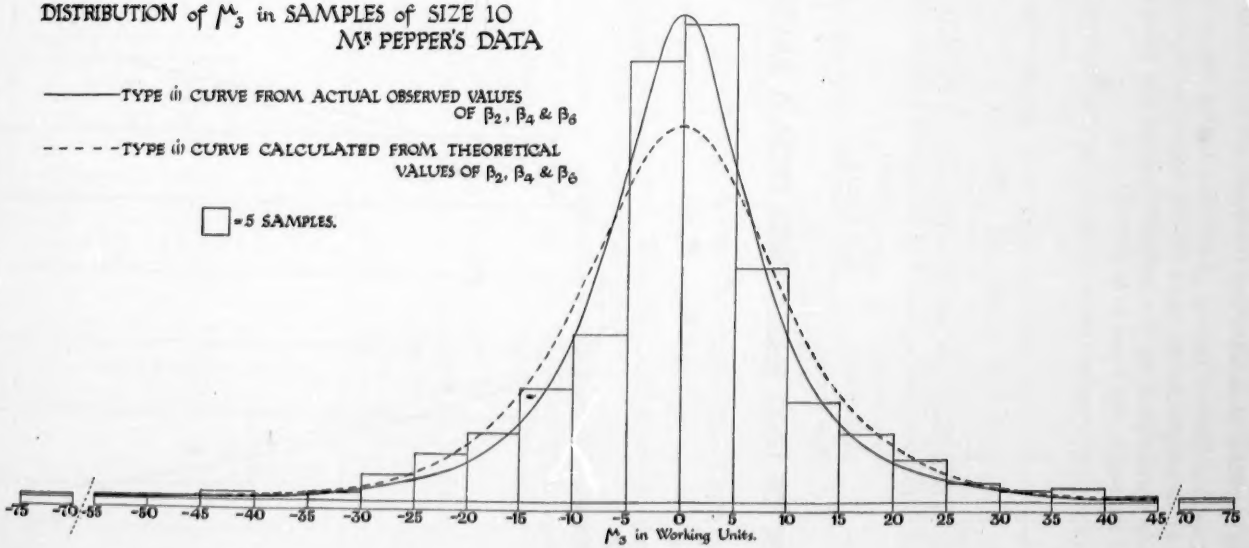

DISTRIBUTION of M_3 in SAMPLES of SIZE 10
Mr PEPPER'S DATA
———TYPE (i) CURVE FROM ACTUAL OBSERVED VALUES OF β_2, β_4 & β_6
- - - - -TYPE (i) CURVE CALCULATED FROM THEORETICAL VALUES OF β_2, β_4 & β_6
□=5 SAMPLES.
-75 -70 -55 -50 -45 -40 -35 -30 -25 -20 -15 -10 -5 0 5 10 15 20 25 30 35 40 45 70 75
M_3 in Working Units.

Diagram XXII.

(*d*) *American Artillery Experience.*

1000 shots were fired from a battery gun in the United States at a target 52 feet long and 11 feet high. All the shots hit the target, and their distribution in the eleven horizontal strips of one foot each, the point of aim being the central horizontal line of the target, was found to be:

No of strip	1	2	3	4	5	6	7	8	9	10	11
Frequency	1	4	10	89	190	212	204	193	79	16	2

I found the mean to be a distance of ·482 from the centre of the sixth strip. The constants referred to the mean and corrected by Sheppard's formulae gave

$$\sigma = 1{\cdot}5499, \qquad \beta_4 = 10{\cdot}9184,$$
$$\beta_1 = \cdot 0008, \qquad \beta_6 = 69{\cdot}2137.$$
$$\beta_2 = 2{\cdot}5261,$$

DISTRIBUTION OF 1000 SHOTS FIRED FROM A BATTERY GUN.

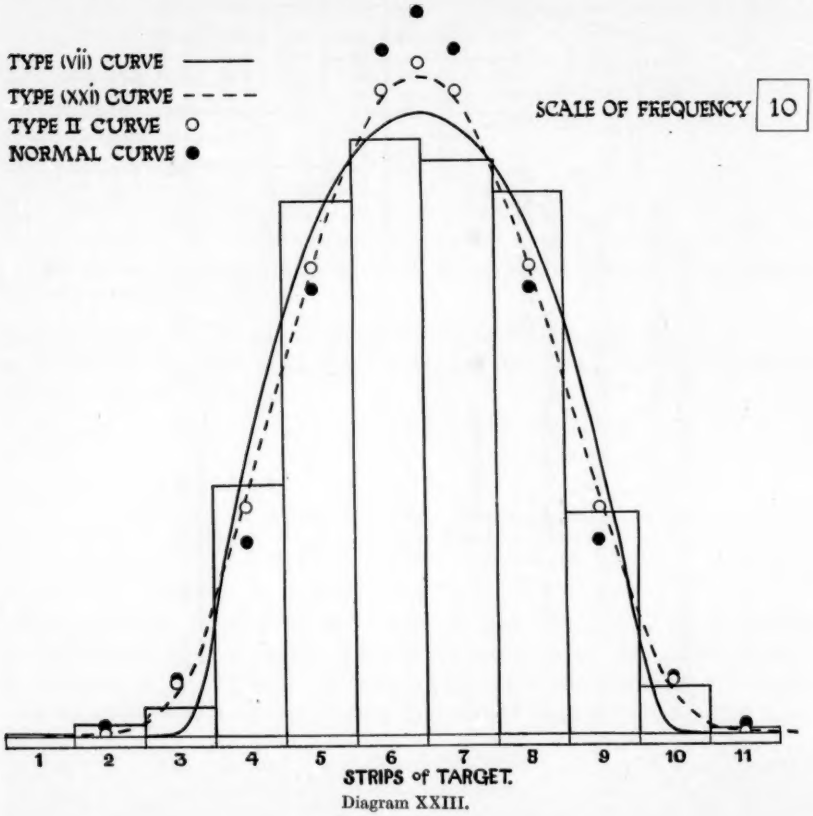


Diagram XXIII.

(i) The ordinates of the transition curves for $\beta_2 = 2{\cdot}5261$ and $\beta_4 = 10{\cdot}9184$ are

$$\begin{aligned} \phi_0 &= 58{\cdot}089, & \phi_6'' &= 68{\cdot}352, \\ \phi_2 &= 33{\cdot}191, & \phi_8' &= 66{\cdot}051, \\ \phi_4 &= 63{\cdot}1822, & \phi_8'' &= 55{\cdot}889. \\ \phi_6' &= 118{\cdot}490, \end{aligned}$$

Since $\phi_6' >$ observed $\phi_6 > \phi_6''$ the point corresponding to the observed β's falls into a heterotypic zone.

I found the ordinates of the Type (vii) curve for

$$\beta_2 = 2{\cdot}526, \quad \beta_4 = 10{\cdot}918, \quad \beta_6 = 61{\cdot}352,$$

but although a better fit is obtained than that given by Pearson Type II curve, the fit is not good. I also tried to fit a sixth order Pearson curve, viz.

$$y = y_0 e^{-\frac{m}{(p-x/\sigma)^2}},$$

which corresponds to the different equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{-4m\,(x/\sigma^2)}{(p - x^2/\sigma^2)^3}.$$

p is given by the cubic*

$$p^3(\beta_2 - 3) + 6p^2\beta_2 - 3p(7\beta_4 - 5\beta_2^2) + 9\beta_6 - 7\beta_2\beta_4 = 0$$

and m by the equation $\quad p^3 - 9p^2 + 15p\beta_2 - 7\beta_4 = 4m.$

Here again the fit is not good. Diagram XXIII shows the above distribution and the ordinates of the various curves plotted.

Note. As a matter of interest I now considered the different types along the line $\beta_4 = 10{\cdot}92$ in the plane $\beta_2 = 2{\cdot}53$. The Table XXX below gives all the types discussed. Diagram XXIV (p. 190) shows some of the various shapes obtained.

TABLE XXX.

β_6	Zone†	Signs of					Type
		m_1	$q_1 p_1$	$q_1 + p_1$	$(q_1 - p_1)^2$	$m_1^2 - (q_1 - p_1)^2$	
30	below $F_2=0$	+	−	+	+	+	(ii)
33·19	$F_2=0$	+	−	0	+	+	(ix)
50	(F_2, F_8'')	+	−	−	+	+	(i)
55·89	$F_8''=0$	+	−	−	+	0	(xiii)
57	(F_8'', F_0)	+	−	−	+	−	(i)
58·09	$F_0=0$	+	0	−	+	−	(xi)
60	(F_0, F_4)	+	+	−	+	−	(iii)
63·18	$F_4=0$	0	+	−	+	−	(viii)
64	(F_4, F_8')	−	+	−	+	−	(vi)
66·05	$F_8'=0$	−	+	−	+	0	(xv)
67	(F_8', F_6'')	−	+	−	+	+	(vi)
68·36	$F_6''=0$	−	+	−	0	+	(vii)
70	(F_6'', F_6')	−	+	−	−	+	(*H*)
118·49	$F_6'=0$	−	+	−	0	+	(vii)
140	above $F_0'=0$	−	+	−	+	+	(vi)

* In our case the cubic has only one real root.

† I denote for shortness by $F_x=0$, on the curve $F_x=0$ and by $(F_x, F_y)=0$, between the curves $F_x=0$ and $F_y=0$. *H* marks that the curve lies in a heterotypic region.

FREQUENCY CURVES for $\beta_2 = 2{\cdot}53$ and $\beta_4 = 10{\cdot}92$, and varying β_6

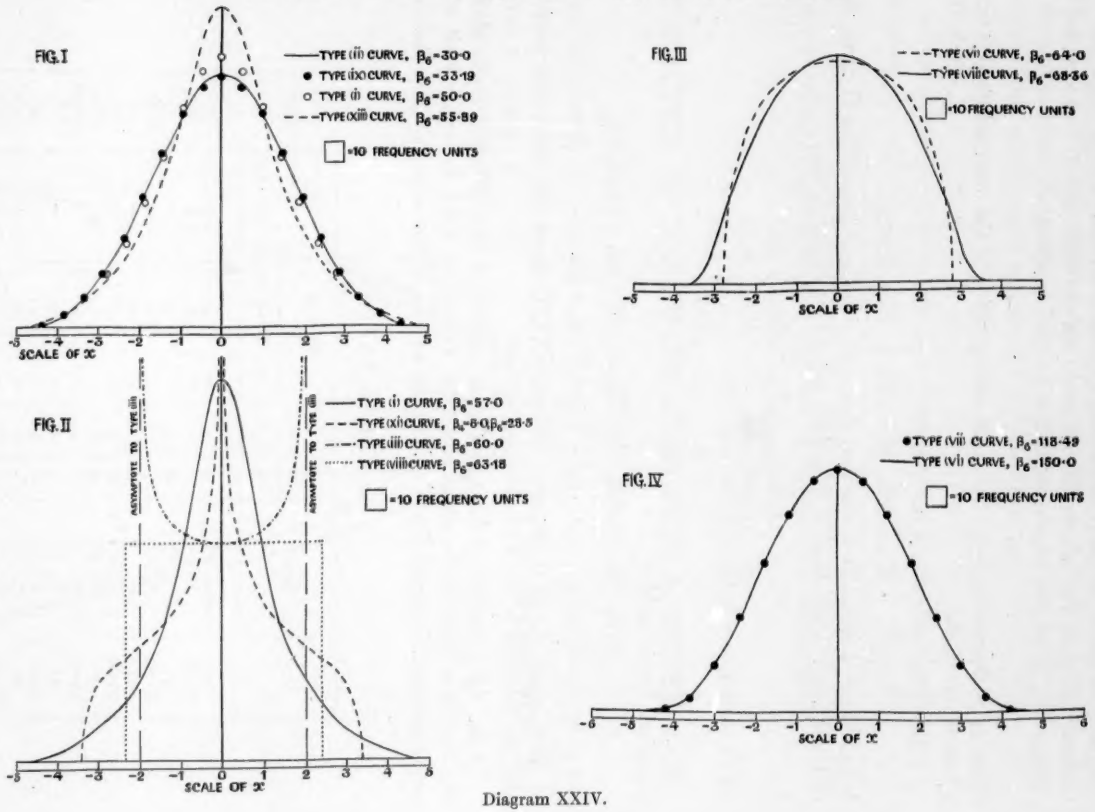


FIG. I
TYPE (ii) CURVE, $\beta_6 = 30{\cdot}0$
TYPE (ix) CURVE, $\beta_6 = 33{\cdot}19$
TYPE (ii) CURVE, $\beta_6 = 50{\cdot}0$
TYPE (xiii) CURVE, $\beta_6 = 55{\cdot}89$
= 10 FREQUENCY UNITS
SCALE OF x
FIG. II
ASYMPTOTE TO TYPE (iii)
ASYMPTOTE TO TYPE (viii)
TYPE (i) CURVE, $\beta_6 = 57{\cdot}0$
TYPE (xi) CURVE, $\beta_6 = 8{\cdot}0, \beta_6 = 28{\cdot}5$
TYPE (iii) CURVE, $\beta_6 = 60{\cdot}0$
TYPE (viii) CURVE, $\beta_6 = 63{\cdot}18$
= 10 FREQUENCY UNITS
SCALE OF x
FIG. III
TYPE (vi) CURVE, $\beta_6 = 64{\cdot}0$
TYPE (viii) CURVE, $\beta_6 = 68{\cdot}36$
= 10 FREQUENCY UNITS
SCALE OF x
FIG. IV
TYPE (vii) CURVE, $\beta_6 = 118{\cdot}42$
TYPE (v) CURVE, $\beta_6 = 150{\cdot}0$
= 10 FREQUENCY UNITS
SCALE OF x

Diagram XXIV.

(*e*) *Errors of the Right Ascensions of Stars**.

Bessel provides the following data:

Limits	Actual Errors	Limits	Actual Errors
0—0·1	114	0·5—0·6	6
·1—	84	·6—	3
·2—	53	·7—	1
·3—	24	·8—	1
·4—	14	·9—	0

The table appears to add the frequencies in the positive and negative groups and to give their sums only. We find the constants

$$\mu_2 = \cdot0513, \quad \mu_4 = \cdot0101, \quad \mu_6 = \cdot00361, \quad \mu_8 = \cdot00174,$$

$$\beta_2 = 3\cdot843, \quad \beta_4 = 26\cdot81, \quad \beta_6 = 252\cdot2.$$

Hence $$r = \left(\beta_4 - \frac{5\beta_2^2}{6-\beta_2}\right) \Big/ \frac{1}{7}\left(\frac{5\beta_2^2}{6-\beta_2} - \beta_2^2\right) = -2\cdot672,$$

$$\Phi = \left(\beta_6 - \frac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2}\right) \Big/ \left\{\frac{25\beta_2^3(\beta_2+8)}{9(6-\beta_2)^2} - \beta_2^3\right\} = -\cdot433.$$

Let us now interpolate linearly into the tables of the Φ_8' ordinates for $\beta_2 = 3\cdot843$ and $r = -2\cdot672$.

Values of Φ_8'.

r	$\beta_2 = 3\cdot5$	$\beta_2 = 3\cdot843$	$\beta_2 = 4$
−2·0	−·352	—	−·382
−2·672	−·451	−·482	−·488
−3·0	−·499	—	−·540

We find $\Phi_8' = -\cdot482$, approximately. But the observed value of Φ_8' is $-\cdot433$ and therefore $\beta_6 > \Phi_8' > \phi_0$, since by inspection of the tables for $\beta_2 = 3\cdot5$ and $\beta_2 = 4$, $\phi_8' > \phi_0$. Accordingly, referring to Diagram I, Fig. III, we see that we must fit the Type (i) curve. Note that we could have reached this result by an inspection of the contour diagrams.

Now $(q-p) = -(q_1+p_1) = 19\cdot206$, therefore $q = 21\cdot107$,

$qp = -q_1p_1 = 40\cdot125$, „ $p = 1\cdot901$,

$m = m_1 = 39\cdot264$, „ $k' = 1\cdot7065$.

* *Astronomische Nachrichten*, Bd. 15, Nr. 358—359.

The Type (i) curve accordingly is $y = y_0 \left(\frac{1 - u^2/21{\cdot}107}{1 + u^2/1{\cdot}901}\right)^{1{\cdot}7065}$, $y_0 = 124{\cdot}0$ and the range is from $-1{\cdot}040$ to $+1{\cdot}040$.

Pearson's Type VII and the normal curves are respectively

$$y = 114{\cdot}145\,(1 + 2{\cdot}140\,x^2)^{6{\cdot}058},$$

and

$$y = 264{\cdot}99 \left(\frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(4{\cdot}4165\,x)^2}\right).$$

The ordinates are plotted in Diagram XXV.

Using six groups the following values of χ^2 were found and hence P by the goodness of fit test:

Type of curve	(i)	VII	Normal
χ^2	·669	1·348	5·517
P	·975	·923	·359

Both the Type (i) and the Type VII give a better fit than the normal curve. There is not much to choose between the Type (i) and the Type VII curves.

THE DISTRIBUTION OF THE RIGHT ASCENSION OF A STAR.
300 OBSERVATIONS MADE BY BESSEL.

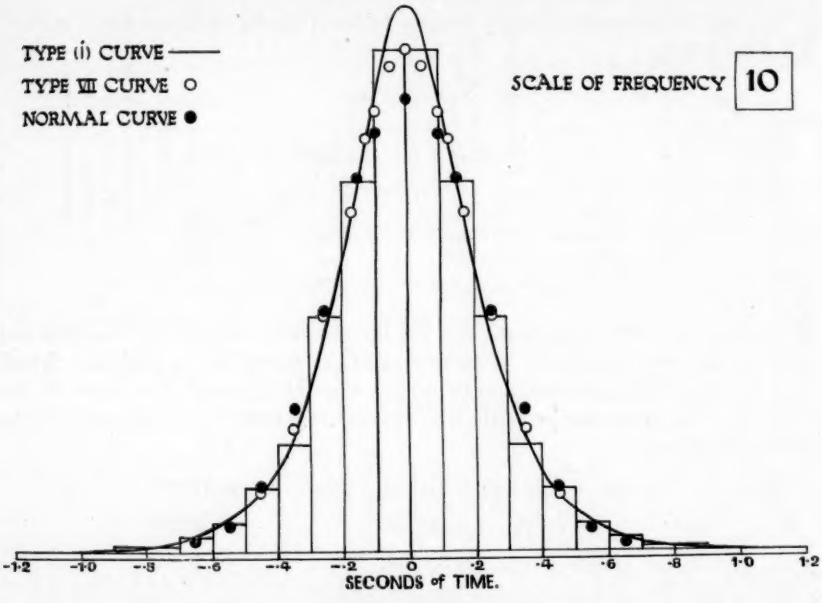


Diagram XXV.

(f) W. Palin Elderton's Data.

The following example is taken from Elderton's "Frequency Curves and Correlation":

Central Value	Observations	Type II	Type A	Edgeworth	Normal
−2·0	11	14	15	16	20
−1·5	116	109	106	106	95
−1·0	274	286	284	285	270
− ·5	451	433	437	436	456
+ ·5	432	433	437	436	456
+1·0	267	285	283	284	270
+1·5	116	109	106	106	95
+2·0	16	14	15	16	20

There is no statement of what this distribution describes, and I have taken the groups as shown above. The constants are

$\mu_1' = -\cdot00149$, $\mu_2 = 1\cdot829$, $\beta_1 = \cdot00237$, $\beta_2 = 2\cdot548$, $\beta_4 = 9\cdot122$, and $\beta_6 = 39\cdot50$.

The corresponding auxiliary co-ordinates are $r = -\cdot682$ and $\Phi = -\cdot049$.

By inspection of the Diagram XV we see that the observed values of r and Φ fall into the zone between the curves of section of $F_0 = 0$ and $F_2 = 0$. Hence, from Diagram I, Type (i) must be used. We find

$$p = 4\cdot926, \quad q = 6\cdot321, \quad k' = 1\cdot330, \quad y_0 = 472\cdot14.$$

The required curve is $$y = y_0 \left(\frac{1 - \frac{x^2}{14\cdot1123}}{1 + \frac{x^2}{9\cdot0103}} \right)^{1\cdot3302}.$$

Note the range is from $-3\cdot76$ to $+3\cdot76$. We find the following frequencies:

Central Value	Observed Frequency	Frequency from Type (i)
−2·0	11	14
−1·5	116	111
−1·0	274	278
− ·5	451	438
+ ·5	432	438
+1·0	267	278
+1·5	116	111
+2·0	16	14

Although we have few groups the Type (i) curve gives a very good fit. All the other types, except the normal curve, fit equally well. Note that in this case the corresponding r and Φ co-ordinates are $r = -\cdot68$ and $\Phi = -\cdot04$, as compared to $r = 0$ and $\phi = \cdot09$ for Pearson's Type II curve.

(g) Distribution in Samples of the Mean of an Array as determined by the Regression Line (K. Pearson, *Proc. Roy. Soc.*, A, Vol. 112, 1926).

Let $\breve{y}_x$, m_1, m_2, σ_1, σ_2, ρ refer to parent populations,

and $\bar{y}_x$, $\bar{x}$, $\bar{y}$, Σ_1, Σ_2, r refer to the sample of size M.

Hence
$$\breve{y}_x = m_2 + \rho\frac{\sigma_2}{\sigma_1}(x - m_1),$$

and
$$\bar{y}_x = \bar{y} + r\frac{\Sigma_2}{\Sigma_1}(x - \bar{x}).$$

The moments of the distribution for $\bar{y}_x' = \bar{y}_x - \breve{y}_x$ are

$$\mu_2 = \frac{\sigma_2^2(1-\rho^2)\phi^2}{M-3}, \text{ where } \phi^2 = 1 - \frac{2}{M} + \frac{(x-m_1)^2}{\sigma_1^2},$$

$$\mu_4 = \frac{3\sigma_2^4(1-\rho^2)^2}{(M-3)(M-5)}\left\{\phi^4 - \frac{2}{M}\left(1-\frac{2}{M}\right)\right\},$$

$$\mu_6 = \frac{15\sigma_2^6(1-\rho^2)^3}{(M-3)(M-5)(M-7)}\left\{\phi^6 - \frac{6}{M}\left(1-\frac{2}{M}\right)\phi^2 + \frac{8}{M^2}\left(1-\frac{2}{M}\right)\right\},$$

$$\mu_8 = \frac{105\sigma_2^8(1-\rho^2)^4}{(M-3)(M-5)(M-7)(M-9)} \times \left\{\phi^8 - \phi^4\left(1-\frac{2}{M}\right)\frac{12}{M} + \phi^2\left(1-\frac{2}{M}\right)\frac{32}{M^2} + \left(1-\frac{2}{M}\right)\left(1-\frac{6}{M}\right)\frac{12}{M^2}\right\}.$$

Hence

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3(M-3)}{(M-5)}\left\{1 - \frac{2}{M}\left(1-\frac{2}{M}\right)\frac{1}{\phi^4}\right\},$$

$$\beta_4 = \frac{15(M-3)^2}{(M-5)(M-7)}\left\{1 - \frac{1}{\phi^4}\frac{6}{M}\left(1-\frac{2}{M}\right) + \frac{8}{M^2}\left(1-\frac{2}{M}\right)\frac{1}{\phi^8}\right\},$$

$$\beta_6 = \frac{105(M-3)^3}{(M-5)(M-7)(M-9)} \times \left\{1 - \frac{12}{M}\left(1-\frac{2}{M}\right)\frac{1}{\phi^4} + \frac{32}{M^2}\left(1-\frac{2}{M}\right)\frac{1}{\phi^6} + \frac{12}{M^2}\left(1-\frac{2}{M}\right)\left(1-\frac{6}{M}\right)\frac{1}{\phi^8}\right\}.$$

(1) Take $M = 50$. From the equations above we find the following table*:

$x - m_1$	β_2	β_4	β_6	r	r_{VII}	Φ	Φ_{VII}
$\frac{1}{2}\sigma_1$	3·051	15·806	119·339	·024	0	·3768	·1874
σ_1	3·102	16·618	129·431	·016	0	·2486	·2051
$2\sigma_1$	3·128	17·045	136·792	·003	0	·2538	·2146

* r_{VII} and Φ_{VII} are the r and Φ co-ordinates of the point corresponding to the Type VII curve.

(2) Take $M = 25$.

$x - m_1$	β_2	β_4	β_6	r	r_{VII}	Φ	Φ_{VII}
$\frac{1}{3}\sigma_1$	3·072	16·195	144·875	·091	0	·5460	·1946
$3\sigma_1$	3·298	20·208	193·408	·068	0	·3326	·2964

(3) For larger samples we are still closer to the point corresponding to Pearson's Type VII curve. For very large samples the curve of distribution of the mean of the arrays follows a normal curve.

We are so close to the point $(0, \Phi_{VII})$ that in all cases the Type VII curve can excellently describe the distribution of the mean of the arrays. We expect this since

$$\beta_2 = \frac{3(M-3)}{M-5},$$

$$\beta_4 = \frac{15(M-3)^2}{(M-5)(M-7)},$$

$$\beta_6 = \frac{105(M-3)^3}{(M-5)(M-7)(M-9)},$$

satisfy $\beta_4 = \frac{5\beta_2^2}{6-\beta_2}$ and $\beta_6 = \frac{35\beta_2^3}{(6-\beta_2)(9-2\beta_2)}$.

Conclusions.

I have shown that the fourth order symmetrical Pearson curves give a better graduation than the second order Pearson curves or the normal curve, and that the improvement in fit, as judged by the χ^2 test, justifies the extra amount of labour involved in finding the higher betas.

I have also succeeded in finding frequency curves whose higher beta-coefficients are positive, in the cases $\beta_2 > 6$ and $4{\cdot}5 < \beta_2 < 6$.

Finally, I have discussed the $\beta_2, \beta_4, \beta_6$-space and the types of curve at different points, but the question of finding the type of curve corresponding to the heterotypic area must be left unanswered. This problem demands an analysis of the sixth order Pearson curves and a discussion of β_8 and β_{10} will be necessary. We cannot expect any advantage from such a system of curves.

In conclusion I would like to thank Professor Karl Pearson for suggesting this problem to me and to express my appreciation of his continual advice and criticism. I wish also to thank Miss M. Kirby for the very capable way in which she has prepared my diagrams.

A BIOMETRIC STUDY OF THE "FLATNESS" OF THE FACIAL SKELETON IN MAN.

By T. L. WOO, Ph.D., in conjunction with G. M. MORANT, D.Sc.

(1) *Introduction.* It has long been recognised that differences in the degree of "flatness" of the face, or facial skeleton, are sufficiently great to be of considerable importance in aiding the discrimination of the racial varieties of modern man. These differences have usually been taken for granted, and it is only in recent years that any attempt has been made to give them quantitative expression. The research described in the present paper was undertaken with the purpose of extending these quantitative methods and applying them to more adequate cranial material than any previously considered. No one appears to have made any serious attempt to define exactly what is meant by the "flatness" of the facial skeleton, but this feature of the cranium has been customarily referred to when discussing measurements of two different kinds. Those of the first are chiefly angular measurements taken in the "median sagittal" plane and the more important of them deal with prognathism and are modifications of Camper's facial angle. The greater the degree of prognathism the less "flat" the face is said to be. Measurements of the projection of the "median sagittal" section of the nasal bones have been dealt with in a similar way by A. Lüthy* and G. von Bonin†, the latter having obtained his data from sagittal type contours. We are not concerned in the present paper with such measurements which are confined to the "median sagittal" plane. Those of the second kind referred to are designed to give estimates, not of antero-posterior projection, but of transverse flattening. One may consider this feature at a number of different and convenient horizons and it is clear that a considerable number of measurements would be needed to give any appreciation of the transverse flattening of the whole facial skeleton. C. de Mérejkowsky seems to have been the first to make a useful contribution to this subject. In a paper which appeared in 1882‡ he described a new instrument—now known as a simometer—designed to measure the projection of the nasal bones. The means found for a number of racial series of crania suggested that his method would be a valuable one, but it was not adopted by craniometricians. Thirty years later a study of negro skulls was issued in *Biometrika*§ in which special attention was paid to the same character. The

* "Die vertikale Gesichtsprofilierung und das Problem der Schädelhorizontalen." *Archiv für Anthropologie*, N.F. Bd. XI. 1912, pp. 1—87.

† *Biometrika*, Vol. XXIII. 1931, pp. 260—261.

‡ "Sur un nouveau caractère anthropologique." *Bulletins de la Société d'Anthropologie de Paris*, 3ième série, t. v. pp. 293—304.

§ "A Study of the Negro Skull with special reference to the Congo and Gaboon Crania." By R. Crewdson Benington. Prepared for press by Karl Pearson. Vol. VIII. 1912, pp. 292—339; see pp. 315—320.

new measurements taken for this purpose, in addition to those defined by de Mérejkowsky, were the chord from dacryon to dacryon and the minimum arc between the same points, so no special instrument was required other than a simometer. A paper published in the same journal in the following year* provided a detailed study in man and the anthropoid apes of the same measurements and also of the subtense from the dacryal chord, the simotic and dacryal subtenses being found with the simometer. Cranial series representing 16 races were dealt with, but several of these were admittedly too small to give reliable means. The dacryal and simotic measurements have been provided for the numerous cranial series described in later volumes of *Biometrika*, and the material available for them is now quite extensive. Their omission from all other recent craniometric studies may be largely attributed to the fact that the simometer is a somewhat complicated instrument which would need to be made to order. Other estimates of the transverse flattening of the facial skeleton—including in that the supraciliary region as Broca did—have been obtained from measurements of horizontal type contours. The section represented is one through the glabella parallel to the Frankfort horizontal plane and indices measuring its curvature anterior to the temporal lines have been given by G. M. Morant for several races†. More recently Davidson Black has defined certain angular measurements to measure the same feature‡. These are found with a goniometer or from the horizontal section on which the projections of certain facial points are marked: they are only available for Chinese series.

The writers' purpose was to make a fairly comprehensive survey of the degree of transverse flattening of the facial skeleton considered as a whole in a sample of races from all parts of the world. The use of any projective methods would have made it impossible to record a sufficiently large number of measurements in the time available, and accordingly they decided to use a simple pair of co-ordinate calipers only. Four pairs of measurements, defined in the next Section, were taken to give four indices, of which all but one are new. The simotic index, which was included, was first defined by de Mérejkowsky, and a considerable amount of material for it has accumulated since. Ryley, Bell and Pearson concluded that, for purposes of racial comparison, "a measure of the mesodacryal index is likely to be of as much value as the determination of the simotic index, possibly of greater value," but nevertheless we chose the latter rather than the former to measure the prominence of the nasal bridge because it can be determined on a far larger number of museum specimens. The lacrymal bones are frequently missing, and it is then impossible to locate the dacrya with precision.

(2) *Definitions of Measurements.* The object of the measurements used in this study was to provide accurate estimates of the transverse flattening of the facial skeleton at different horizons. This was effected by measuring directly, with the

* "A Study of the Nasal Bridge in the Anthropoid Apes and its Relationship to the Nasal Bridge in Man." By Kathleen V. Ryley and Julia Bell, assisted by Karl Pearson. Vol. IX. 1913, pp. 391—445.

† *Annals of Eugenics*, Vol. II. 1927, pp. 355—356.

‡ *Palaeontologia Sinica*, Series D, Vol. VII. 1928, pp. 13—14.

aid of a special pair of calipers, the subtense of a particular "median" point from a corresponding chord, of which the terminals are the same point on the right and left sides, and also the length of the chord. The index obtained by expressing the subtense as a percentage of the chord provides a measure of the character with which we are concerned, and four different indices were found, whenever possible, for each skull. The readings required could be taken rapidly and accurately with a pair of co-ordinate calipers. This instrument consists of a bar, on which a scale is inscribed, with three arms which always remain mutually parallel to one another and perpendicular to the bar such that their tips, which are brought into contact with the object measured, are always co-planar with the scale. One of the arms is fixed to the zero end of the bar, and another of the same length can be moved along it so that these two are used in precisely the same way as the arms of an ordinary pair of rectangular calipers. The third is attached between the other two, and it can also be moved along the bar, the attachment permitting a movement so that the setting point of the arm can be shifted towards or away from the bar. This middle arm bears a scale on which the reading is zero when the tips of the three arms are in a straight line. When the middle arm is moved towards or away from the bar the divergence of its tip from that straight line—i.e. the subtense—can be read. In our case the tips of the outside arms were first brought into contact with the extremities of the transverse chord, and then the middle arm was moved until its tip was in contact with the appropriate "median" point, a screw enabling it to be fixed in that position. On removing the calipers both the length of the chord and the subtense of the "median" point from it were recorded. Another reading might have been taken at the same time. This is the distance of the foot of the perpendicular—i.e. the "working edge" of the middle arm—from one or other extremity of the chord which would have given an appreciation of asymmetry. The measurement in question was not recorded, but for two series the direct distance of the "median" point from the extremity of the chord was found on either side. The form of co-ordinate calipers made by P. Hermann, Richenbach und Sohn, of Zürich, was found to be unsuitable for taking the measurements devised for two reasons: the middle arm is too broad to make it possible to obtain the adjustment required in the case of narrow nasal bones, and the outside arms are not long enough to enable other subtenses as large as some needed to be found. Accordingly a similar instrument was made to order for Dr Woo by Messrs W. F. Stanley and Co. The breadth of the middle arm was reduced to 2·5 mm., and the outside arms were made 10 cm. long. All the measurements needed could be readily taken with these calipers on all the skulls examined.

The measurements are:

(1) *IOW.* The internal bi-orbital breadth between the points, right and left, where the fronto-malar sutures cross the outer margins of the orbits. The point used, which is Martin's *fronto-malare orbitale*, can almost always be found precisely if the suture be open. If the orbital margin be blunt, it is advisable to mark it first as a pencil line. The measurement is No. 43 (1) in Martin's list.

(1 *a*) Sub. *IOW*. The subtense of the nasion from the chord *IOW*.

(1 *b*) The frontal index of facial flatness is defined to be 100 Sub. *IOW*/*IOW*.

(2) *SC*. The simotic chord which is the minimum horizontal breadth of the two nasal bones, so that the extremities of this chord are points on the naso-maxillary sutures.

(2 *a*) *SS*. The simotic subtense found by first marking as a pencil line the "ridge" of the nasal bones, without necessarily following the inter-nasal suture, and then finding the minimum subtense from this line to the simotic chord. When found with co-ordinate calipers this may not correspond exactly with the simotic subtense found with the aid of de Mérejkowsky's simometer. A comparison between the two measurements found for the same series is made below.

(2 *b*) The simotic index is defined to be 100 *SS*/*SC*.

(3) *MOW*. The mid-orbital breadth between the points, right and left, where the malar-maxillary sutures cross the lower margins of the orbits. It is generally advisable to mark the lower margins of the orbits as pencil lines before locating the points.

(3 *a*) Sub. *MOW*. The subtense from the "tip" of the nasal bones, which is accepted to be the lowest point on the inter-nasal suture (i.e. Martin's *rhinion*), to the chord *MOW*. This was found to be an unsatisfactory measurement owing to the fact that the lower parts of the nasal bones are often defective, even on skulls which have the remainder of the facial skeleton quite complete. It is sometimes difficult to decide whether the region in question is intact or not, and in all doubtful cases the measurement was omitted.

(3 *b*) The rhinal index is defined to be 100 Sub. *MOW*/*MOW*.

(4) *GB*. The chord between the points, right and left, on the zygomatico-maxillary sutures which are lowest with regard to the Frankfort horizontal plane. The positions of these points can be estimated with sufficient accuracy without orientating the cranium exactly. If the inferior extremity of the suture is a short length lying parallel to the horizontal plane, the anterior point on it is the one accepted. The point used is Martin's *Zygomaxillare* and the chord (No. 46 in his list) is one usually given in the descriptions of racial series.

(4 *a*) Sub. *GB*. The subtense of the alveolar point from *GB*. The true alveolar point was used, i.e. the inferior point on the process between the cavities of the central incisors, and not the prosthion.

(4 *b*) The premaxillary index of facial flatness is defined to be 100 Sub. *GB*/*GB*.

The first three of these indices are of a straightforward character, and in each case the three points involved usually lie in a plane which is approximately parallel to the Frankfort horizontal plane. For the first the points all lie on the inferior borders of processes of the frontal bone; the second measures the flatness of the bridge of the nose and it is confined to the nasal bones, while the third is of a less simple type since it involves the nasal and maxillary bones. The points used in determining the fourth index all lie on the premaxillary bones,

but the "median" point—the alveolar point—is always decidedly inferior to the extremities of the transverse chord. The subtense will obviously be affected by differences in the height of the premaxillary region, though the index may be supposed to give primarily a measurement of sub-nasal prognathism. We should anticipate that the latter would be highly correlated with a measurement of the total prognathism of the face.

Of the absolute measurements defined above, *IOW*, *SS*, *SC* and *GB* have been provided for a number of cranial series, and all except the first were previously available for several of the series measured for the purpose of this study. Hence, comparisons can be made to test the reliability of the same measurement taken on the same specimens by two different observers, and this is particularly necessary in the case of the simotic subtense as two different instruments, and slightly different definitions have been used to determine it. De Mérejkowsky describes how he measured the character in question with his simometer. The two arms of the instrument are placed so that their tips coincide with the extremities of the minimum breadth of the two nasal bones, and then, to bring the tip of the subtense arm into position, "on l'appuie sur le dos du nez là où celui-ci atteint son minimum d'élévation; les trois points de repaire (*sic!*) sont donc faciles à trouver et complètement fixes*." But the last statement does not seem to be true since "le dos du nez" is an inexact term and yet it forms an essential part of the definition, as it is only with reference to it that the *minimum d'élévation* can have any meaning. We must suppose that de Mérejkowsky was imagining a "median" line, which might be traced in pencil on the nasal bones, such that it would indicate their "ridge," or the line marking the join of the right and left sides of the surface formed by the two nasal bones together. On a symmetrical cranium this would coincide with the inter-nasal suture, but the direction of the suture was not considered, presumably, in determining the "ridge" if the nasal bones were asymmetrical. The subtense would then be the *minimum* from this line to the simotic chord, and having fixed its position it would also be the maximum subtense of the section of the nasal bones defined by the three points of contact of the instrument. This method of measuring the simotic subtense appears to have been that followed by Ryley, Bell and Pearson, and by later workers in the Biometric Laboratory who used the simometer. Dr Woo used co-ordinate calipers in place of that instrument, but followed the same convention in making the point of subtense of the subtense arm one on the "median" ridge of the nasal bridge. When the compass arms of the simometer are fixed, as to the simotic chord of a particular cranium, the pivot of the instrument, which is equidistant from their tips, is also a fixed point. The middle, or subtense, arm is then free to rotate round that pivot in such a way that the tips of the three arms and the pivot are always co-planar. It is adjusted so that its tip touches the point on the accepted transverse section of the nasal bones which is nearest to the pivot. The simometer is then removed from the cranium and, while the direction of the

* *Loc. cit.* p. 298.

subtense arm relative to the compass arms is maintained, its tip is moved until it is in line with the tips of the compass arms. The subtense measured is the displacement of the tip of the subtense arm*. If the section in which the simometer measurements are taken is asymmetrical, the subtense may not be perpendicular to the chord and the point of maximum subtense may not coincide with that found by using the co-ordinate calipers. It is clear that the simometer subtense will always be equal to or greater than the subtense found with co-ordinate calipers, and we must ask whether the readings found with the two different instruments are sufficiently close to be considered comparable for practical purposes. To decide the point Dr Woo found the nasal bridge measurements with co-ordinate calipers in the case of a series—the Kerma Egyptian measured by Miss Collett—for which the measurements had previously been recorded by using a simometer. Comparisons are made in the following table, the constants relating to differences having been found from the distributions of differences:

Character	Sex	Measured by	Mean	Mean Difference (C. – W.)	S.D. of Differences
SS	♂	Collett with simometer Woo with co-ordinate calipers	4·006 (94) 3·843 (94)	+ ·162 ± ·032	·464 ± ·023
	♀	Collett with simometer Woo with co-ordinate calipers	3·125 (68) 3·147 (68)	− ·022 ± ·029	·353 ± ·020
SC	♂	Collett Woo	10·81 (94) 10·36 (94)	+ ·450 ± ·034	·489 ± ·024
	♀	Collett Woo	10·22 (68) 9·99 (68)	+ ·229 ± ·040	·488 ± ·028
100 *SS*/*SC*	♂	Collett (simometer subtense) Woo (co-ordinate calipers subtense)	37·24 (94) 36·57 (94)	+ ·681 ± ·276	3·962 ± ·195
	♀	Collett (simometer subtense) Woo (co-ordinate calipers subtense)	30·58 (68) 31·76 (68)	− 1·184 ± ·232	2·834 ± ·164
GB	♂	Collett Woo	95·27 (97) 95·10 (97)	+ ·167 ± ·038	·549 ± ·027
	♀	Collett Woo	90·83 (82) 90·71 (82)	+ ·120 ± ·044	·587 ± ·031

* The simometer is rather a complicated instrument and the reader who is not acquainted with it may find it difficult to follow our remarks without refering to de Mérejkowsky's description and figure in the paper cited.

The mean difference between the male simotic subtenses (*SS*) is seen to be small but quite significant, the mean found by using the simometer being greater than that given by the co-ordinate calipers. The difference between the female means, however, is of the opposite sign though quite insignificant. But some significant differences are also found for the simotic chords (*SC*) and facial breadths (*GB*), though in these cases the two observers followed precisely the same definitions.

For the male simotic index the difference is actually insignificant, while for the female it is quite significant but of the opposite sign to that which would have been expected. We may conclude that the differences between the two methods of determining the simotic subtense only introduce errors which are of the same order as those due to the personal equations of two different workers following the same definition*, and accordingly we have supposed that the simotic measurements found with the co-ordinate calipers may be compared without correction with those previously found with the simometer. For all the series dealt with (except the Kerma Egyptian) which had any of the measurements *IOW*, *SS*, *SC*, 100 *SS*/*SC* and *GB* provided in previously published papers, those values were accepted to give the constants provided in the present paper. It may be observed that none of the differences shown in the table on p. 201 is large enough to vitiate racial comparisons, except possibly the differences for the simotic chord, since they are very small compared with interracial differences. For the simotic index, for example, a difference of about one unit is shewn between the means found by the different observers for the same series, but the range of the mean male simotic indices for different series is from 20·7 to 53·1.

In the next place, we may consider the few measurements collected for the purpose of examining asymmetry. For two male series the direct distances of each median point from the extremities of the corresponding chord were recorded and a comparison of these bilateral measurements is made in the table on p. 203.

Only one of the differences of means can be considered significant and if there be any differentiation of the bilateral measurements considered it is evidently of a small order which would only be revealed by examining larger series.

(3) *Description of the Material Measured.* The object of the present paper is to test the racial significance of the new measurements of facial flatness defined in the previous section. In order to serve this purpose it was necessary to measure as large numbers as possible of crania representing as many races as possible from all parts of the world. The observer (T.L.W.) was obliged to limit his material to that contained in the principal English museums and one Dutch museum. He is greatly indebted to the curators of these institutions for affording him every facility for

* In the paper by Ryley, Bell and Pearson (*loc. cit.* pp. 396—398) comparisons are made between various mean simotic measurements found by the same observer who repeated them on the same series of crania at different times. The largest differences shown are 0·10 for *SS*, 0·36 for *SC* and 1·0 for 100*SS*/*SC*. These are of the same order as, but rather less than, the largest differences shown in the table above for the corresponding measurements.

pursuing this study and particularly to Miss M. L. Tildesley (Royal College of Surgeons), Dr L. H. Dudley Buxton (Oxford), Dr W. L. H. Duckworth (Cambridge), Mr W. P. Pycraft (British Museum), Professor J. A. J. Barge (Leiden) and, finally, to Professor Karl Pearson (Biometric Laboratory). The number of adult crania actually measured was 4266 male, 1630 female, the sexes accepted being in general those provided in previously published papers or museum catalogues, though in the case of a few series the writers had to assign sexes to the specimens themselves. No immature specimens were dealt with. Whenever possible long series representing single racial types were chosen and a number of earlier studies—nearly all of which have appeared in *Biometrika*—dealing with their characters were consulted as a guide to the groups which could be considered racially homogeneous. In the case of a number

Constants	Series		Frontal	Simotic	Rhinal	Premaxillary
Means	Punjabi	*R*	52·5 ± ·18 (81)	6·47 ± ·08 (81)	36·5 ± ·41 (33)	57·9 ± ·25 (67)
		L	52·2 ± ·19 (81)	6·44 ± ·09 (81)	35·8 ± ·44 (33)	58·2 ± ·27 (67)
	Hindu (Madras Presidency)	*R*	51·4 ± ·25 (49)	6·05 ± ·10 (46)	35·2 ± ·74 (10)	57·6 ± ·36 (48)
		L	51·3 ± ·23 (49)	5·95 ± ·11 (46)	34·6 ± ·80 (10)	58·0 ± ·34 (48)
Standard Deviations	Punjabi	*R*	2·44 ± ·13	1·12 ± ·06	3·45 ± ·29	3·08 ± ·18
		L	2·47 ± ·13	1·16 ± ·06	3·77 ± ·31	3·24 ± ·19
	Hindu (Madras Presidency)	*R*	2·61 ± ·18	1·10 ± ·07	—	3·69 ± ·25
		L	2·44 ± ·17	1·13 ± ·08	—	3·44 ± ·24
Correlations (r_{RL})	Punjabi		·828 ± ·024	·826 ± ·024	·949 ± ·012	·917 ± ·013
	Hindu		·953 ± ·009	·942 ± ·011	—	·863 ± ·025
Differences of Means ($\Delta_{M_R - M_L}$)	Punjabi		+ ·3 ± ·108*	+ ·03 ± ·051	+ ·7 ± ·140	− ·3 ± ·107
	Hindu		+ ·1 ± ·076	+ ·10 ± ·038	+ ·6 ± ·255†	− ·4 ± ·183

of races, however, it was only possible to arrive at means based on sufficiently large numbers by pooling the measurements of short series in different museums. Or means of larger series might be pooled if there was sufficient reason to believe that they represented the same racial type. The test used in such cases was based only on the four indices of facial flatness. If two or more series were thought, from the description of their origins, to represent the same racial type, then their mean indices were calculated and if no significant differences were found between them the series were pooled to give the constants finally used. This practice is only of provisional value, of course. Two samples of skulls may show one character, or group of

* The probable error of the difference of means is given by the formula

$$\cdot 67449/\sqrt{n} \times \sqrt{\sigma_R^2 + \sigma_L^2 - 2r_{RL}\sigma_R\sigma_L}.$$

† This approximate probable error was found by using the σ's and r given for the Punjabi series.

characters, identically the same while differing most significantly in other ways. Several of the series finally used are probably less homogeneous racially than any which could legitimately be used for purposes of rigid racial classification, but they may still be of value in a preliminary study such as the present. All the continental areas are represented, but the races used do not form a random sample from all races of the world. There are several British series, but few long series from other parts of Europe; the races of Africa and Oceania are fairly well represented; the American and Western Asiatic series are particularly meagre, while there are none from Central Asia and Siberia, but India and the Orient are better represented than any other regions. The means are given of a few series which are so short that no use can be made of them. It may be possible to pool these with measurements collected later from other sources.

The museums in which the cranial specimens were measured are: The Museum of the Royal College of Surgeons, London (R.C.S.); the British Museum, Natural History (B.M.); the Biometric Laboratory, University College, London (Biom. Lab.); the University Anatomical Museum at Cambridge (Cambridge); the Oxford University Museum (Oxford); and the Museum of the "Anatomisches Laboratorium," Leiden, Holland (Leiden). The abbreviations given in the brackets are used to denote them in this paper. *Bm.* denotes this *Journal* and *J.A.I.* the *Journal of the Anthropological Institute.*

A. European Series.

I. *English: Long Barrow.* Cambridge 18 ♂. The skulls in Thurnam's collection came chiefly from different localities in Wiltshire. They are of the Neolithic period.

II. *Romano-British.* Cambridge 11 ♂, 3 ♀; Oxford 81 ♂, 43 ♀. The specimens at Cambridge came from a single cemetery at White Horse Hill, Berkshire, and those at Oxford came from a number of counties, mainly from Oxfordshire, Berkshire and Wiltshire.

III. *Anglo-Saxon.* (*a*) Cambridge 13 ♂, 8 ♀; Oxford 41 ♂, 19 ♀. These skulls came from a considerable number of cemeteries in different parts of England. (*b*) Cambridge 42 ♂, 20 ♀, only the complete skulls being measured. This series was recently excavated by Mr Lethbridge at Burwell, Cambs., from a cemetery of seventh century date. In order to justify the pooling of the measurements of the above two series the male means and standard deviations for the four indices dealt with were calculated separately so that the significance of the differences of the means for these characters between the two series could be satisfactorily tested. The values are given in the Table at the top of p. 205.

It is evident that all the differences between the means are quite insignificant. This justifies the provisional pooling of the two series and the combined means, which have been used for comparative purposes in this paper, are given in Table I.

IV. *English: Mediaeval.* Oxford 43 ♂, 11 ♀. These crania came chiefly from Abingdon, Berkshire, and Sittingbourne, Kent. A few unspecified specimens are also included.

Indices	Frontal		Simotic	
	Mean	σ	Mean	σ
Series (*a*)	18·6 ± ·19 (52)	2·07 ± ·14	52·3 ± 1·06 (44)	10·41 ± ·75
Series (*b*)	18·5 ± ·22 (40)	2·06 ± ·16	52·4 ± ·87 (33)	7·54 ± ·63
(*a*) − (*b*)	+0·1 ± ·29	—	−0·1 ± 1·38	—
	Rhinal		Premaxillary	
Series (*a*)	45·4 ± ·85 (16)	5·04 ± ·60	35·1 ± ·36 (42)	3·50 ± ·26
Series (*b*)	45·2 ± 1·39 (6)*	—	35·1 ± ·47 (34)	4·10 ± ·34
(*a*) − (*b*)	+0·2 ± 1·63	—	0·0 ± ·60	—

V. *English: Farringdon Street.* Biom. Lab. 81 ♂, 76 ♀. This series came from a London churchyard used in the seventeenth century. See *Bm.* Vol. XVIII. pp. 1—55, 1926.

VI. *English: Spitalfields.* Biom. Lab. 248 ♂, 94 ♀. This series came from a burial-ground at Spitalfields Market, London, and it is of mediaeval or Roman date. See *Bm.* Vol. XXIII. pp. 191—248, 1931.

VII. *Modern Irish.* R.C.S. 19, Cambridge 8, Oxford 38 (soldiers): all ♂. Of these only 15 are from known localities of the Irish Free State. The majority of the crania, including 38 soldiers and 5 seamen, are from unknown localities in Ireland.

VIII. *Modern Scottish.* R.C.S. 19, Cambridge 3, Oxford 13 (soldiers): all ♂. Twenty skulls came from known localities of Scotland and others are unspecified.

IX. *British (Soldiers).* (*a*) English soldiers: Oxford 31. Their native places are unknown. (*b*) British soldiers: Oxford 82. No detailed information relating to the origin of these specimens is available.

Indices	Frontal		Simotic	
	Mean	σ	Mean	σ
Series (*a*)	19·1 ± ·29 (31)	2·35 ± ·20	51·5 ± 1·65 (30)	13·36 ± 1·16
Series (*b*)	18·9 ± ·18 (82)	2·46 ± ·13	53·1 ± ·97 (81)	12·84 ± ·68
(*a*) − (*b*)	+ ·2 ± ·34	—	−1·6 ± 1·91	—
	Rhinal		Premaxillary	
Series (*a*)	43·2 ± ·90 (19)	5·82 ± ·64	35·6 ± ·32 (27)	2·45 ± ·23
Series (*b*)	44·7 ± ·49 (65)	5·89 ± ·35	36·3 ± ·24 (74)	3·03 ± ·17
(*a*) − (*b*)	−1·5 ± 1·03	—	− ·7 ± ·36	—

* The probable error of the mean of the rhinal index for the (*b*) series is based on the standard deviation of the same index for the (*a*) series.

The mean measurements of the above two series have been pooled to give those in Table I (p. 222). It is evident from the figures shown above that the mean differences for the four indices between the two series are too small to be significant. It is, therefore, legitimate to pool them.

The last three series VII—IX above are probably less homogeneous racially than any other British series used. The means derived from them can only be used provisionally.

X. *North German.* R.C.S. 28, Cambridge 4: all ♂. The specimens came principally from different provinces of Prussia—Brandenburg, Holstein, Westphalia, Hessen and Saxony being represented.

XI. *French.* R.C.S. 46 ♂, 19 ♀. These skulls came mainly from the catacombs of Paris, though a few are from unknown localities. They belong to the post-Merovingian period.

XII. *Italian.* R.C.S. 102 ♂, 33 ♀; Cambridge 1 ♀. These specimens, which previously belonged to G. Nicolucci, came from 12 divisions in the northern and central parts of Italy. A few from the islands of Sicily and Sardinia were excluded. They are all of modern date.

XIII. *Norwegian.* R.C.S. 19 ♂. These specimens came from a Seamen's Hospital at Greenwich, so they are presumably the skulls of seamen. One individual is stated to have come from Bergen, but the home districts of the others are not known.

XIV. *Swedish.* R.C.S. 30 ♂. Of these, 9 crania came from the neighbourhood of Stockholm, and the others are only marked as seamen.

XV. *Finnish.* R.C.S. 22 ♂, Cambridge 1 ♂. Ten of these crania came from various localities of Finland, and 13 from a Seamen's Hospital collection are recorded as Russian Finns. As the series is a short one, no useful comparisons could be made between the two groups.

XVI. *Dutch.* R.C.S. 29 ♂, 10 ♀; Leiden 22 ♂. Two-thirds of these specimens came from different known provinces of Holland, viz. Guelderland, Utrecht, Friesland, Groningen, Zeeland and North Holland: the others are unspecified.

XVII. *Ancient Greek.* R.C.S. 3 ♂, Oxford 9 ♂. These skulls came from Athens, Attica and the Ionian Islands, and they are dated approximately between 500 and 200 B.C.

XVIII. *Modern Greek.* R.C.S. 18 ♂, Oxford 15 ♂. The localities from which these specimens came are: 8 from the Greek mainland, 12 from the Ionian Islands, 4 from the Aegean Islands, and 9 from unknown districts of Greece.

XIX. *Czech.* R.C.S. 23 ♂, 10 ♀. These specimens came from Bohemia, Slovakia and Moravia. The central part of Moravia is better represented than the other regions, so the series may be loosely called a Czech one.

XX. *Russian.* R.C.S. 20 ♂, Leiden 16 ♂. Of these specimens, 22 came from various scattered localities in Russia, and others are marked as "unknown" in locality.

European races, apart from British, are much less satisfactorily represented than are those of the other continental areas except America. This is chiefly due to the fact that there are few European skulls other than British in English museums. The pooling of several short series from various localities of the same country had to be effected in several cases in order to obtain sufficiently large samples.

B. African Series.

I. *Guanche.* R.C.S. 16 ♂, 19 ♀; Cambridge 5 ♂, 4 ♀; Leiden 4 ♂, 5 ♀. Nearly all these specimens came from Teneriffe.

II. *Nubian.* Biom. Lab. 69 ♂, 61 ♀. This series was excavated by the Archaeological Survey of Nubia in 1907—1908 from sites immediately south of Assuan. It consists only of New Empire (17th—20th Dynasty) crania.

III. *Egyptian: Badari.* Biom. Lab. 36 ♂, 22 ♀. See *Bm.* Vol. XIX. pp. 110—150, 1927. This series is believed to be the earliest of predynastic date that has yet been discovered in Egypt. It has been shown that the earlier Egyptian series bear a closer resemblance to negro types than do the later series from the same country.

IV. *Egyptian: Sedment.* Biom. Lab. 40 ♂, 30 ♀. See *Bm.* Vol. XXII. pp. 65—92, 1930. This series came from a cemetery of 9th Dynasty date at Gebel Sedment, which is about 70 miles South of Cairo.

V. *Egyptian: Kerma.* Biom. Lab. 117 ♂, 97 ♀. These skulls represent the Egyptian population of Kerma (Nubia) in the 12th and 13th Dynasties. See *Bm.* Vol. XXV. pp. 254—284, 1933.

VI. *Negro: Nigeria.* (*a*) Crania from North Nigeria. R.C.S. 6 ♂, 4 ♀; Cambridge 13 ♂, 3 ♀. These skulls came from various parts of North Nigeria. (*b*) Crania from South Nigeria. R.C.S. 34 ♂, 12 ♀. These skulls came from several provinces of the country, and the majority of them are recorded as representing the Ibibio and Ekoi tribes of the Calabar region. The male means of these two series for the four indices were calculated separately and are shown below. The probable errors of the North Nigeria means were found by using the South Nigeria standard deviations.

Indices	Frontal	Simotic	Rhinal	Premaxillary
(*a*) N. Nigeria (Mean)	17·7 ± ·29 (18)	30·7 ± 1·68 (17)	30·2 ± 1·16 (11)	36·2 ± ·56 (17)
(*b*) S. Nigeria (Mean)	17·6 ± ·21 (34)	28·9 ± 1·22 (32)	29·5 ± 1·11 (12)	36·5 ± ·43 (29)
S. Nigeria (σ)	1·82 ± ·15	10·24 ± ·86	5·70 ± ·79	34·2 ± ·30
(*a*) − (*b*)	+·1 ± ·36	+1·8 ± 2·08	+·7 ± 1·61	−·3 ± ·71

It is obvious that none of the differences between the corresponding pairs of indices should be regarded as significant. The pooled means in Table I (p. 222) were used for comparative purposes.

VII. *Gaboon Negro.* (*a*) Gaboon, 1864. B.M. 50 ♂, 44 ♀. See *Bm.* Vol. VIII. pp. 292—339, 1912. This series was collected by Du Chaillu in Fernand Vaz. (*b*) Gaboon, 1880. B.M. 18 ♂, 19 ♀. See *ibid.* This series came from the same locality, and it was collected by the same person. It has been shown that series (*a*) and (*b*) bear a very close resemblance to one another when all the usual cranial characters are compared. (*c*) Gaboon, collected at various times. R.C.S. 12 ♂, 6 ♀. These specimens came from the neighbourhood of Fernand Vaz and the Gaboon river. The means of the above three male series for the four indices are shown in the following table.

Indices	Frontal	Simotic	Rhinal	Premaxillary
(*a*) Gaboon, 1864 (Mean)	17·5 ±·21 (46)	30·9 ± ·91 (45)	28·9 ± ·41 (32)	38·5 ±·37 (40)
(*b*) Gaboon, 1880 (Mean)	17·5 ±·34*(18)	29·0 ±1·52 (16)	29·2 ± ·64 (13)	38·5 ±·64 (13)
(*c*) Gaboon, R.C.S. (Mean)	18·3 ±·42*(12)	28·3 ±1·92 (10)	28·7 ±1·16 (4)	39·3 ±·73 (10)
Gaboon, 1864 (σ)	2·13±·15	9·01± ·64	3·43± ·29	3·42±·26
(*a*) − (*b*)	0±·40	+1·9 ±1·77	− ·3 ± ·76	0±·74
(*a*) − (*c*)	− ·8 ±·47	+2·6 ±2·12	+ ·2 ±1·23	− ·8±·82
(*b*) − (*c*)	− ·8 ±·54	+ ·7 ±2·45	+·5 ±1·32	− ·8±·97

It is evident that the differences between all possible pairs of means are quite insignificant, and the combined means given in Table I may therefore justifiably be employed for comparison in this paper.

VIII. *Angoni.* B.M. 25 ♂. See *J.A.I.* Vol. XXVIII. pp. 55—94, 1898. The specimens were obtained from M'ponda's Town at the south end of Lake Nyassa. The donor, Sir H. H. Johnston, describes them as being the crania "of a slightly mixed negro race mainly belonging to the Anyanja stock with a slight Zulu intermixture. It is possible, however, that one or more of them may be Yaos mixed with Arab blood."

IX. *Ashanti.* R.C.S. 12 ♂; Cambridge 2 ♂; Oxford 20 ♂, 31 ♀. Of these, 4 male specimens came from Kumasi, and 3 males were of Ashanti warriors killed in war in 1873—74. Others are merely marked as having been obtained from districts on the West coast of Africa, and no further information is available.

X. *Teita Negro.* Biom. Lab. 57 ♂, 65 ♀. See *Bm.* Vol. XXIII. pp. 271—314, 1931. The Teita, or Wa-Teita, are a small tribe of Bantu-speaking negroes living in the South of Kenya Colony.

XI. *Zulu.* R.C.S. 17 ♂. Most of these skulls are of Zulu warriors. They came from two battle-fields on the Zulu border, Isandhlwana and Ulundi, where fighting took place in 1856 and 1879 respectively.

* The probable errors of the four indices for series (*b*) and (*c*) were obtained by using the corresponding standard deviations of series (*a*).

XII. *Congo Negro.* R.C.S. 58♂, 30♀; Leiden 2♂, 1♀. See *Bm.* Vol. VIII. pp. 292—339, 1912. The majority of these specimens represent the Batetela tribe who live at about 24° 20′ E., 4° 51′ S. near the Lubefu River. A few came from the neighbourhood of the Congo basin and the Upper Congo regions.

XIII. *Bushman.* R.C.S. 9♂, 14♀; Cambridge 2♂; B.M. 1♂; Oxford 11♂, 3♀. Of these, there are 4 males and 1 female from the Cape of Good Hope, 1 male from the Transvaal, 1 male from Namaqualand, and 1 female from the Orange Free State, while the localities from which the others came are unknown.

XIV. *Hottentot.* R.C.S. 2♂, 6♀; Oxford 7♂, 2♀. Of these, 3 males and 1 female are known to have come from Namaqualand, Middleburg, the Cape of Good Hope and the Cape of St Francis respectively, while the others are from unknown localities.

XV. *Kaffir.* R.C.S. 22♂, 7♀; Cambridge 5♂, 2♀; Oxford 25♂, 1♀; B.M. 3♂. Various tribes of Kaffirs are represented in this series, but most of the specimens represent Bantu negroes inhabiting the South and South-East of Africa. Six male specimens are stated to have been obtained from native burial-grounds during the Kaffir War in 1847.

C. American Series.

I. *Eskimo.* R.C.S. 33♂, 17♀; Cambridge 4♂, 1♀; Oxford 11♂, 11♀; B.M. 4♂, 1♀. Of these, 31 males and 18 females are marked as having come from various parts of Greenland and the neighbouring islands; 12 males and 8 females are from several islands situated to the North of Baffin Land and from the North-East coast of Labrador. A few came from unknown localities. All the specimens have been pooled together to make up a sufficient number for the purpose of statistical enquiry.

II. *Peruvian.* This is one of longest series which was measured for the present purpose. In order to examine local differences in the characters dealt with, the material may be first divided into two groups: (*a*) Specimens from Pacasmayo. R.C.S. 20♂, 13♀; Cambridge 47♂, 22♀; Oxford 4♂, 4♀. These mostly belong to Consul Hutchinson's Collection. (*b*) Specimens from other localities. R.C.S. 37♂, 28♀; Cambridge 10♂, 5♀; Oxford 55♂, 42♀. These were mostly purchased from Consul Hutchinson. The localities from which the majority of these skulls came are: Cerro del Oro (in the Canete Valley, 100 miles South-East of Callao), Ancon, Huacho (near Callao) and Santo, all situated in the Western part of Peru. The first two named are better represented than the others. The means of the four indices for the two male series are given in the table at the top of p. 210.

From the figures shown it is apparent that the mean differences are all too small to be considered significant except in the case of the simotic index for which the difference is 2·6 times its probable error. It was considered that for the present purpose all these Peruvians might be pooled together although a significant difference occurs in the second index. It should be noted here that among the Peruvian skulls there are many which show a fronto-occipital deformation. It may

Indices	Frontal		Simotic	
Constants	Mean	σ	Mean	σ
(*a*) Peruvians from Pacasmayo ...	17·2 ± ·13 (69)	1·61 ± ·09	40·3 ± ·77 (67)	9·30 ± ·54
(*b*) Peruvians from other localities	17·2 ± ·13 (102)	1·98 ± ·09	42·9 ± ·64 (99)	9·49 ± ·45
(*a*) − (*b*)	0 ± ·19		− 2·6 ± 1·00	
	Rhinal		Premaxillary	
(*a*) Peruvians from Pacasmayo ...	40·0 ± ·84 (16)	4·99 ± ·59	34·6 ± ·26 (61)	2·96 ± ·18
(*b*) Peruvians from other localities	39·1 ± ·61 (27)	4·72 ± ·43	34·2 ± ·22 (82)	2·98 ± ·16
(*a*) − (*b*)	+ ·9 ± 1·04		+ ·4 ± ·34	

be asked: Does such a deformation affect their facial measurements? This question can be easily answered by making a comparison of the means for each character from two sets of crania, i.e. crania recorded as having a slight or marked artificial deformation in the frontal region and crania preserved in a natural state. Let us take the male Peruvian specimens at Oxford for illustration as the state of deformation is individually recorded in the catalogue of the Museum there. The indices for the two groups just mentioned were calculated separately and the standard deviations based on the total male Peruvian crania were employed to obtain the probable errors. The constants are:

Indices	Frontal	Simotic	Rhinal	Premaxillary
(*a*) Normal crania	16·7 ± ·19 (43)	43·9 ± 1·00 (41)	38·3 ± ·84 (15)	33·0 ± ·36 (32)
(*b*) Deformed crania	17·0 ± ·31 (16)	41·2 ± 1·66 (15)	40·0 ± 1·33 (6)	34·1 ± ·56 (13)
(*a*) − (*b*)	− ·3 ± ·36	+ 2·7 ± 1·94	− 1·7 ± 1·58	− 1·1 ± ·66

It will be seen that the ratio of the difference to its probable error is in no case greater than 1·7. This indicates that our facial measurements are not affected to a significant degree by the artificial deformation practised by the Peruvians as far as can be ascertained from the short series available.

III. *North American Indian.* R.C.S. 38 ♂, 26 ♀; Cambridge 8 ♂. This series represents a miscellaneous collection of North American Indians derived chiefly from Vancouver Island and the West. The different tribes are represented by such small numbers that there is no alternative to pooling all the available material; it gives means which are of small permanent value. Some specimens show a certain degree of frontal deformation, but their facial skeletons were presumably not distorted to any appreciable extent.

IV. *Patagonian.* R.C.S. 21♂. These specimens came from various districts in the neighbourhood of the South-East coast of Patagonia, North of Tierra del Fuego. They represent several tribes of native Indians.

V. *Fuegian.* R.C.S. 12♂. These came from the South-West coast of Tierra del Fuego.

D. Oceanic Series.

I. *Maori: New Zealand.* R.C.S. 44♂, 13♀; Cambridge 10♂, 5♀; Oxford 29♂, 6♀; B.M. 6♂, 1♀. Of these, 30 males and 9 females came from the North Island, principally from Auckland, and a few came from the South Island, but there are 60 specimens, male and female, the original localities of which are not specified.

II. *Moriori.* R.C.S. 33♂, 21♀; Cambridge 3♂, 1♀; Oxford 6♂, 4♀; B.M. 1♂, 1♀. These specimens came from various parts of the Chatham Islands and it may be assumed that they form a racially homogeneous sample.

III. *New British.* Cambridge 39♂, 24♀. The specimens were collected by Mr A. Willey in the Island of New Britain.

IV. *Islanders of the Marquesas.* R.C.S. 22♂, 16♀. The series came from five main Islands, viz. Nukahiva, Fatuhiva, Hivaoa, Uahuka and Hiau. .

V. *Easter Islanders.* R.C.S. 35♂, 16♀; Leiden 11♂, 5♀. A detailed description of this material is given in Dr von Bonin's paper: *Bm.* Vol. xxiii. pp. 249—270, 1931.

VI. *Kanaka: Sandwich Islands.* R.C.S. 52♂, 54♀; Cambridge 15♂, 6♀. "Kanakas" is the name applied by some anthropologists to the natives of the Polynesian Islands generally. In the native language of the Sandwich Islands "Kanaka" means "man." Here it is restricted to the inhabitants of the Sandwich Islands alone. More than 85 per cent. of the crania of this series came from the Island of Oahu (Woahoo) and the remainder from the Island of Hawaii (Owyhee).

VII. *New Caledonian.* R.C.S. 18♂. Of these, 10 specimens came from the mainland of New Caledonia, 4 from the Island of Pines, a small island South-East of New Caledonia, and 4 from the Island of Lifu, the largest of the Loyalty Islands, near New Caledonia.

VIII. *Solomon Islanders.* R.C.S. 21♂, Cambridge 6♂. Among these 3 specimens are known to have come from the Island of San Cristoval, 1 from Guadalcanao and 1 from Isabel. The others are unspecified and only stated as being natives of the Solomon Islands

IX. *Islanders of New Hebrides.* R.C.S. 36♂. Of these, 24 crania came from various Islands of the New Hebrides ranging from the Cherry (Anuda) Island in the North to the island of Tanna in the South. These include Eromanga, Ambrym, Faté, Api and the Banks Islands. The remaining 12 crania came from the Island of Mallicollo, and these are all remarkable for the depression of the frontal region. It has not yet been ascertained whether this is a natural conformation, or whether

it is due to artificial deformation in infancy. The evidence derived from the Peruvian series leads us to hope that the facial measurments of any deformed crania in the present series would not have been affected appreciably.

X. *Fijian.* R.C.S. 30 ♂, 13 ♀. Out of these, 18 males and 9 females came from various burial-grounds on the Island of Viti Levu, the largest island of the Fiji group. They are mostly of the coastal tribes. Six males and 4 females came from the Island of Ovalau and 6 males from the Island of Vanua-Valavo.

XI. *Papuan.* Cambridge 79 ♂, 49 ♀. The specimens were collected by the Daniels' Expedition, and they came from more than 10 different districts in the territory of British New Guinea.

XII. *Islanders of the Cook Group.* R.C.S. 19 ♂. Of these, 9 came from the Island of Atiu (20° 0′ S., 158° 5′ W.), 4 from the Island of Mangaia and the remainder from neighbouring small islands.

XIII. *Tasmanian.* R.C.S. 21 ♂, 13 ♀. These specimens were collected by various persons in Tasmania.

XIV—XVI. *Australian.* It was found convenient to group the native Australian skulls measured in the following way:

(*a*) Australians from the Northern Territory. R.C.S. 10 ♂, Oxford 7 ♂, B.M. 1 ♂. These skulls came chiefly from Port Essington, Port Darwin, the region of Van Diemen Gulf and the banks of the Victoria River, so all are from the North-West of the Northern Territory.

(*b*) Australians from Queensland. R.C.S. 25 ♂, 10 ♀; Cambridge 2 ♂, 1 ♀; Oxford 9 ♂, 1 ♀. More than half these specimens came from North Queensland and the remainder from South, East and Central Queensland.

(*c*) Western Australians. R.C.S. 8 ♂, 3 ♀; Cambridge 6 ♂, 6 ♀; B.M. 3 ♂, 2 ♀. These skulls came chiefly from the North-Western district of Western Australia.

(*d*) Australians from New South Wales. R.C.S. 12 ♂, 8 ♀; Cambridge 2 ♂, 2 ♀; Oxford 17 ♂, 8 ♀. These specimens came mainly from Sydney, Port Stephens, Bathurst, and the neighbourhood of the Hunter, Macleay and Clarence Rivers, i.e. from regions close to the Eastern coast of New South Wales.

(*e*) South Australians. R.C.S. 35 ♂, 26 ♀; Cambridge 11 ♂, 5 ♀; Oxford 5 ♂, 4 ♀; B.M. 1 ♂, 2 ♀. Most of these came from Adelaide and the banks of the Murray River, near the South coast of South Australia.

(*f*) Australians from Victoria. R.C.S. 18 ♂, 13 ♀; Cambridge 2 ♂; Oxford 4 ♂, 1 ♀. The majority of these crania came from Melbourne, Geelong, Port Phillip and Port Fairy, on the South coast of Victoria. A few came from Ballarat and the vicinity of the Murray River (Victoria).

We require to know whether there are any significant differences between the characters of the six Australian series. In order to examine this point the means of the male series for the four indices were calculated separately and they are given

below. The differences between all possible pairs of means together with their probable errors are presented in the table below. The standard deviations of the South Australian series were used in computing the probable errors of the means for the Northern Territory, Western Australian, New South Wales and Victoria groups. The differences which exceed 2·5 times their probable errors are in italics. Considering the differences between the means in terms of their corresponding probable errors, which are shown in the second part of the table below, it will at once be noticed that the mean indices of the last four Australian series show no significant differences from one another, but the first two series, i.e. Australians from the Northern Territory and from Queensland, differ quite significantly in one or two characters from one another and also from the last four series mentioned, the single exception being that the Northern Territory series shows no significant differences from the South Australian group. Accordingly, the whole of the Australian specimens measured can be best regrouped in three series, viz. Australians from the Northern Territory (XIV), from Queensland (XV), and from all other parts of the continent (XVI). It may be noticed that the frontal indices show no significant differences. The general characters of Australian crania—according to Morant's recent study

Constants	Series	Frontal	Simotic	Rhinal	Premaxillary
Means	(*a*) Australians from Northern Territory	18·8 ± ·39 (18)	41·9 ± 1·50 (16)	37·2 ± ·93 (12)	41·5 ± ·60 (17)
	(*b*) Australians from Queensland	18·5 ± ·28 (36)	37·5 ± 1·01 (36)	30·1 ± ·67 (18)	39·5 ± ·39 (35)
	(*c*) Western Australians	18·9 ± ·41 (16)	43·8 ± 1·45 (17)	33·4 ± 1·08 (9)	38·9 ± ·75 (11)
	(*d*) Australians from New South Wales	18·5 ± ·30 (30)	43·2 ± 1·09 (30)	33·7 ± ·81 (16)	39·2 ± ·45 (30)
	(*e*) South Australians	18·2 ± ·23 (52)	45·9 ± ·90 (44)	35·0 ± ·78 (17)	40·6 ± ·39 (41)
	(*f*) Australians from Victoria	18·4 ± ·34 (23)	43·3 ± 1·25 (23)	33·3 ± 1·14 (8)	39·8 ± ·54 (21)
Standard Deviations	(*b*) Australians from Queensland	2·53 ± ·21	8·98 ± ·71	4·18 ± ·47	3·41 ± ·28
	(*e*) South Australians	2·45 ± ·16	8·87 ± ·62	4·78 ± ·55	3·68 ± ·27
Differences of Means	(*a*) − (*b*)	+ ·3 ± ·48	+ 4·4 ± 1·80	+ *7·1 ± 1·19*	+ *2·0 ± ·72*
	(*a*) − (*c*)	− ·1 ± ·57	− 1·9 ± 2·08	+ *3·8 ± 1·42*	+ *2·6 ± ·96*
	(*a*) − (*d*)	+ ·3 ± ·49	− 1·3 ± 1·85	+ *3·5 ± 1·23*	+ *2·3 ± ·76*
	(*a*) − (*e*)	+ ·6 ± ·45	− 4·0 ± 1·75	+ 2·2 ± 1·22	+ ·9 ± ·72
	(*a*) − (*f*)	+ ·4 ± ·52	− 1·4 ± 1·95	+ *3·9 ± 1·47*	+ 1·7 ± ·81
	(*b*) − (*c*)	− ·4 ± ·50	− *6·3 ± 1·77*	− *3·3 ± 1·26*	+ ·6 ± ·85
	(*b*) − (*d*)	0 ± ·41	− *5·7 ± 1·49*	− *3·6 ± 1·04*	+ ·3 ± ·60
	(*b*) − (*e*)	+ ·3 ± ·38	− *8·4 ± 1·35*	− *4·9 ± 1·03*	− 1·1 ± ·55
	(*b*) − (*f*)	+ ·1 ± ·45	− *5·8 ± 1·61*	− 3·2 ± 1·32	− ·3 ± ·67
	(*c*) − (*d*)	+ ·4 ± ·51	+ ·6 ± 1·82	− ·3 ± 1·35	− ·3 ± ·87
	(*c*) − (*e*)	+ ·7 ± ·47	− 2·1 ± 1·71	− 1·6 ± 1·33	− 1·7 ± ·84
	(*c*) − (*f*)	+ ·5 ± ·54	+ ·5 ± 1·91	+ ·1 ± 1·57	− ·9 ± ·92
	(*d*) − (*e*)	+ ·3 ± ·38	− 2·7 ± 1·41	− 1·3 ± 1·12	− 1·4 ± ·60
	(*d*) − (*f*)	+ ·1 ± ·46	− ·1 ± 1·66	+ ·4 ± 1·40	− ·6 ± ·71
	(*e*) − (*f*)	− ·2 ± ·41	+ 2·6 ± 1·54	+ 1·7 ± 1·39	+ ·8 ± ·67

(*Bm.* Vol. XIX. pp. 417—440, 1927)—suggest a division into two groups, viz. Northern Australians, on the one hand, and all the remainder, on the other. This accords with the present grouping based on measurements of facial flattening only, except that we have also to divide off a Queensland group. The pooled means of each character for our three Australian series are given in Table I (p. 222).

E. Asiatic Series.

I. *Chinese.* Two series of Chinese crania were measured in the museums of Holland and England:

(*a*) Chinese (Leiden). Leiden 48 ♂, 4 ♀. Of these, 15 males are marked as having come from the provinces of Kwantung and Fukien, South China, 14 males and 2 females from Hong Kong and Batavia (Java), and the remaining crania are unspecified. There is some reason to believe that these unspecified specimens most probably belong to the Southern Chinese and that no Northern elements are involved.

(*b*) Chinese (English museums). R.C.S. 67 ♂, 12 ♀; Cambridge 6 ♂; Oxford 13 ♂, 2 ♀; B.M. 3 ♂. Nearly half of the specimens measured are marked as having come from Canton, Macao, Amoy, Hong Kong and Batavia (Java), and these almost certainly represent the Southern Chinese. Others are marked as criminals, seamen, or merely inscribed as "Chinese," and no further information is available. These unspecified crania, we believe, may be the heads of Southern Chinese as the spellings of their names given on the skulls are most likely Cantonese or Fukienese. The means and variations of the four indices for both male Chinese series are given in the following table.

Indices	Frontal		Simotic	
	Mean	σ	Mean	σ
(*a*) Chinese (Leiden)	15·9 ± ·22 (48)	2·29 ± ·16	32·8 ± ·77 (48)	7·92 ± ·55
(*b*) Chinese (England)	15·8 ± ·16 (88)	2·15 ± ·11	31·1 ± ·86 (89)	11·97 ± ·61
(*a*) − (*b*)	+ ·1 ± ·28	—	+ 1·7 ± 1·15	—
	Rhinal		Premaxillary	
(*a*) Chinese (Leiden)	31·9 ± ·52 (43)	5·08 ± ·37	34·8 ± ·32 (44)	3·14 ± ·23
(*b*) Chinese (England)	31·2 ± ·43 (66)	5·19 ± ·31	33·9 ± ·28 (77)	3·57 ± ·19
(*a*) − (*b*)	+ ·7 ± ·68	—	+ ·9 ± ·42	—

Again we find that for all the characters with which we are dealing the two series differ insignificantly from one another, although these samples were collected in various parts of South China and its neighbouring islands. This accords with the belief that the Chinese race is homogeneous in its physical type.

II. *Burmese.* Two series of Burmese skulls were measured.

(*a*) Burmese *A* Type. Biom. Lab. 44 ♂, 39 ♀. A southern series from the neighbourhood of Moulmein was divided into three groups of which the Burmese *A*, according to Miss Tildesley's study (*Bm.* Vol. XIII. pp. 176—262, 1921), is supposed to be true Burman.

(*b*) Burmese in general. R.C.S. 32 ♂; Oxford 10 ♂. The specimens in these museums are only marked "from Burma." Their localities are mostly unspecified with the exception of 7 cases marked "from Rangoon or Pegu." All skulls known to represent the primitive native races of Burma were omitted.

The differences of the mean indices between the two male series are tested in the usual way.

Indices	Frontal		Simotic	
	Mean	σ	Mean	σ
(*a*) Burmese *A* ...	15·8 ± ·28 (42)	2·72 ± ·20	32·9 ± ·97 (41)	9·24 ± ·69
(*b*) Burmese in general	16·2 ± ·16 (42)	1·57 ± ·12	33·3 ± ·97 (41)	9·25 ± ·69
(*a*) − (*b*)	− ·4 ± ·33	—	− ·4 ± 1·38	—
	Rhinal		Premaxillary	
(*a*) Burmese *A* ...	28·7 ± ·52 (21)	3·53 ± ·37	32·7 ± ·35 (34)	3·00 ± ·25
(*b*) Burmese in general	29·8 ± ·43 (36)	3·80 ± ·30	33·6 ± ·32 (36)	2·85 ± ·23
(*a*) − (*b*)	− 1·1 ± ·67	—	− ·9 ± ·47	—

It is seen that so far as the indices of facial flattening are concerned the two series do not differ significantly from one another. It is, therefore, justifiable to pool them for the present comparative purposes.

III. *Japanese.* R.C.S. 9 ♂, Cambridge 1 ♂, Oxford 4 ♂, B.M. 3 ♂. The series is a short one as there are only a few Japanese specimens in English museums. Only three crania are marked as having come from Tokio, Yokohama and Kyoto respectively, the others being from unknown parts of Japan. They are all of modern date.

IV. *Aino.* R.C.S. 7 ♂, B.M. 1 ♂. Of these, 5 came from the island of Yezo (Hokkaido), the homeland of the Aino, 2 came from the Kurile Islands and 1 from an unknown locality.

V. *Tibetan A.* R.C.S. 17 ♂, B.M. 20 ♂. These specimens came from the South-West of the country and are classified by Morant as of the Tibetan *A* Type. See *Bm.* Vol. XIV. pp. 193—260, 1923, and Vol. XVI. pp. 1—105, 1924.

VI. *Tibetan B.* R.C.S. 15 ♂. This series represents soldiers from the Eastern Province of Khams who are said to belong to the Tibetan *B* Type. See Morant, *loc. cit.* 1923.

VII. *Nepalese.* B.M. 48 ♂. These crania came from different parts of Nepal. See *Bm.* Vol. XVI. pp. 1—105, 1924.

VIII. *Malay.* R.C.S. 10 ♂. These specimens were collected in the Malay States.

IX. *Andamanese.* R.C.S. 17 ♂, 16 ♀; Cambridge 14 ♂, 9 ♀. These specimens came mainly from the Great Andaman Islands although several crania are unspecified. South Andaman is better represented than the other islands.

X. *Nicobarese.* Cambridge 14 ♂. These specimens were collected by J. M. Wright in 1923—24.

XI. *Naga.* Biom. Lab, 23 ♂, 15 ♀. These specimens were on loan to the Biometric Laboratory at the time when they were measured (see *Bm.* Vol. XXV. pp. 1—20, 1933). They represent natives of the Naga Hills in Eastern Assam.

XII. *Javanese.* The Javanese skulls measured may be divided into the four groups below.

(*a*) Javanese from Bantam and Batavia. Leiden 55 ♂, 12 ♀. This series came from the provinces of Bantam and Batavia in the Western part of the island. See *Bm.* Vol. XXIII. pp. 52—113, 1931.

(*b*) Javanese from Middle and East Java. Leiden 69 ♂, 15 ♀. These skulls came principally from the Middle and East of the island although a few unspecified skulls are also included. See *ibid.*

(*c*) Javanese from Madura. R.C.S. 10 ♂, Leiden 13 ♂.

(*d*) Javanese in general. R.C.S. 30 ♂, 7 ♀. These skulls came from various parts of Java.

Constants	Series	Frontal	Simotic	Rhinal	Premaxillary
Means	(*a*) Javanese (B.B.)	15·5 ± ·20 (54)	33·4 ± ·92 (53)	30·4 ± ·55 (46)	35·3 ± ·38 (51)
	(*b*) „ (M.E.)	15·9 ± ·17 (69)	32·7 ± ·82 (67)	30·4 ± ·37 (61)	34·1 ± ·32 (67)
	(*c*) „ (Madura)	16·0 ± ·31* (23)	31·6 ± 1·44 (22)	30·7 ± ·71 (21)	35·0 ± ·57 (22)
	(*d*) „ (General)	16·1 ± ·26* (30)	33·4 ± 1·24 (30)	30·5 ± ·73 (20)	33·9 ± ·52 (26)
Standard Deviations	(*a*) Javanese (B.B.)	2·18 ± ·14	9·88 ± ·65	5·54 ± ·39	4·01 ± ·27
	(*b*) „ (M.E.)	2·08 ± ·12	10·13 ± ·58	4·25 ± ·26	3·87 ± ·23
Differences of Means	(*a*) − (*b*)	− ·4 ± ·26	+ ·7 ± 1·23	0 ± ·66	+1·2 ± ·50
	(*a*) − (*c*)	− ·5 ± ·37	+1·8 ± 1·71	− ·3 ± ·90	+ ·3 ± ·68
	(*a*) − (*d*)	− ·6 ± ·33	0 ± 1·54	− ·1 ± ·91	+1·4 ± ·64
	(*b*) − (*c*)	− ·1 ± ·35	+1·1 ± 1·66	− ·3 ± ·80	− ·9 ± ·65
	(*b*) − (*d*)	− ·2 ± ·31	− ·7 ± 1·49	− ·1 ± ·81	+ ·2 ± ·61
	(*c*) − (*d*)	− ·1 ± ·40	−1·8 ± 1·90	+ ·2 ± 1·01	+1·1 ± ·77

* The mean weighted value of the standard deviations of the first two series $\left(\bar{\sigma}=\sqrt{\frac{n_1\sigma_1^2+n_2\sigma_2^2}{n_1+n_2}}\right)$ for each index was used in calculating the probable errors of the last two series.

As the mean differences of the indices between all possible pairs of these series, which are shown in the table on p. 216, are too small to be considered significant, the four Javanese series have justifiably been pooled for the present purpose and the combined means are given in Table I. It has been shown by von Bonin that cranial series from different parts of Java are very similar to one another in type.

XIII. *Sumatran.* R.C.S. 8♂, Leiden 48♂. These crania came from various parts of the country. A large number of them are said to come from the residencies of Palembang and Lampong and from the West coast of Sumatra. Several from Nias and other neighbouring islands have been included.

XIV. *Natives of Celebes.* R.C.S. 9♂, Leiden 43♂. This series is chiefly made up of the skulls of Macassars and Bugis.

XV. *Natives of Sarawak.* R.C.S. 9♂, Cambridge 46♂, 24♀. The specimens preserved in the two museums were all collected by Sir Charles Hose in Borneo in 1899. They represent various tribes belonging to different districts of Sarawak, chiefly those of the Murut group. They were mostly killed by Long Pata, Long Kiputs or other tribes in the Baram district, Sarawak.

XVI. *Dayak.* R.C.S. 13♂, 8♀, Leiden 42♂, 10♀. See *Bm.* Vol. XXIII. pp. 52—113, 1931, for an account of the Leiden series. As all these skulls came from Borneo and are marked as Dayaks, the two series may justifiably be pooled.

XVII. *Tagal.* Leiden 31♂, 19♀. See *ibid.* The bulk of these crania came from Manila, Mindanao and Central Luzon, Philippine Islands. The Tagals are usually classified as one of the brown, non-negrito tribes of the islands.

XVIII. *Aëta.* Leiden 33♂, 14♀. See *ibid.* The Aëtas (Philippine Islands) are found in the interior of Luzon Island and in the islands of Mindoro, Panay and Negros, and also in the North-Eastern part of Mindanao. The people are very short in stature and are usually classed as Negritoes.

XIX. *Singalese.* R.C.S. 24♂, 10♀; Cambridge 1♂, 2♀; Oxford 9♂, 1♀. The majority of these specimens came from Colombo, Ceylon, and a few from different parts of the Western and Central Provinces of Ceylon. The crania of peoples other than Singalese living in the island were excluded.

XX. *Veddah.* R.C.S. 16♂, 10♀; Cambridge 2♂, 1♀; Oxford 3♂, 3♀. These skulls came mainly from the South-Western part of the Eastern Province of Ceylon.

XXI. *Punjabi.* The Indians living in the Province of Punjab are, according to their religions, chiefly Mohammedan and Hindus. It is supposed that the people of the two groups do not inter-marry. The specimens measured represent: (*a*) Mohammedans. R.C.S. 39♂, Cambridge 6♂. This series comprises all crania inscribed as Mohammedan or Moslem from the Punjab. (*b*) Hindus. R.C.S. 42♂, Cambridge 6♂. This series consists of various castes of Hindus among which the low castes are better represented than the others. From the figures shown below it will be seen

that no significant difference is found between any corresponding pair of indices for these two series. They have therefore been pooled.

Series	Frontal		Simotic	
	Mean	σ	Mean	σ
(*a*) Mohammedans	21·2 ± ·28 (45)	2·79 ± ·20	49·6 ± 1·28 (45)	12·72 ± ·90
(*b*) Hindus	20·9 ± ·25 (48)	2·55 ± ·18	49·2 ± 1·26 (47)	12·79 ± ·89
(*a*) − (*b*)	+ ·3 ± ·37	—	+ ·4 ± 1·79	—
	Rhinal		Premaxillary	
(*a*) Mohammedans	43·5 ± ·77 (23)	5·45 ± ·54	36·6 ± ·37 (39)	3·41 ± ·26
(*b*) Hindus	43·6 ± 1·44 (20)	9·55 ± 1·02	35·7 ± ·39 (40)	3·64 ± ·28
(*a*) − (*b*)	− ·1 ± 1·63	—	+ ·9 ± ·54	—

XXII. *Hindu: Madras Presidency.* R.C.S. 49 ♂. These skulls came from the following localities: (*a*) 30 crania from Berhampur in the North-East of the Presidency; (*b*) 6 from the Bellary district, in the Western part of the Presidency; (*c*) 6 from a burial-ground near Madras, and (*d*) 7 from the Salem district in the South of the Presidency. All these specimens have been pooled for the purpose of the present paper.

XXIII. *Hindu: North-East India.* There are three more Hindu series which were measured in different museums. These came from the Provinces of Bengal, Bihar and Orissa, and the United Provinces, all in the North-East of India, so they may be conveniently dealt with together. The origin of these series is described as follows:

(*a*) Bengalese. R.C.S. 88 ♂, 38 ♀; Cambridge 8 ♂, 9 ♀; Oxford 19 ♂, 5 ♀. The majority of these specimens represent various castes of Hindus living in different districts of Bengal. Several crania of Mohammedans which came from the same province are also included.

(*b*) Hindus from Bihar and Orissa. R.C.S. 37 ♂. These skulls represent various castes of Hindus living in the two provinces mentioned. The specimens came largely from the Patna district in the North-West of Bihar.

(*c*) Hindus from the United Provinces. R.C.S. 26 ♂; Cambridge 1 ♂. These specimens represent several castes of Hindus who inhabited the United Provinces. The low castes of Hindus are better represented than the high.

The significance of the differences between the mean indices for these three series and those from the Punjab and Madras (Nos. XXI and XXII above) is tested as usual and the constants are given in the Table on p. 220. The differences which exceed 2·5 times their probable errors are in italics.

From the figures shown in the lower part of the table (p. 220) it may be observed that the Punjabi (XXI) and the Southern Hindus (XXII) differ significantly from one another in a slight degree in the case of the first two indices; also, both these series show slightly significant differences, in one or two characters, from all the three series from North-East India, with the single exception that no significant difference is found between the series from Bihar and Orissa and that from Madras Presidency (XXII and XXIII *b*). The last two represent adjoining regions. The largest difference, which is 4·3 times its probable error, is found for the rhinal index between the Punjabi (XXI) and the Bengalese (XXIII *a*), which represent regions as far apart as any pair. In the case of the last three Hindu series (XXIII *a*—*c*) the differences of the means for all possible pairs of indices are quite insignificant. This shows that the degree of "facial flatness" of these samples from an enormous Indian population is a constant feature as far as can be told from the slender evidence available, and such a result agrees with that of a previous study which was based on the usual cranial measurements of several short Hindu series: see *Bm.* Vol. XX[B]. pp. 294—300, 1928. In view of the comparisons made above, the five East Indian series measured may be justifiably reduced to three groups, viz. the skulls from the Punjab, those from North-East India (XXIII *a*—*c*) and those from Madras Presidency. The grouping seems to be a reasonable one which accords with the geographical positions of the peoples compared.

XXIV. *Dravidian.* R.C.S. 36 ♂. These specimens belong to two short series: (*a*) 18 crania which came chiefly from the Madura district (Madras Presidency), and belong to the Maravar tribe, which is Dravidian; (*b*) 18 crania from the Tanjore district (Madras Presidency), and no detailed information relating to these is available, but the locality suggests that they are most probably of Dravidian origin.

XXV. *Afghan.* R.C.S. 18 ♂. These crania came from various parts of Afghanistan. Six of them are recorded as belonging to the Yusafzai tribes in the Peshawar Valley of this country.

XXVI. *Arab.* R.C.S. 10 ♂. This collection came from scattered parts of the country including Western (the coast of Midian), Southern (Hadramaut) and Eastern (Oman) Arabia.

XXVII. *Natives of Syria and Palestine.* R.C.S. 10 ♂ from Syria, 11 ♂ from Palestine; Cambridge 4 ♂ from Syria. This series came from several cemeteries in various districts of the two countries. Owing to the small number of specimens measured the pooling of these crania from the two neighbouring countries has been considered desirable for the purpose of the present investigation only.

	Indices	Frontal		Simotic		Rhinal		Premaxillary	
	Series	Mean	σ	Mean	σ	Mean	σ	Mean	σ
Means and σ's	XXI. Punjabi	21·0 ± ·20 (93)	2·87 ± ·14	49·4 ± ·91 (92)	12·94 ± ·64	43·5 ± ·79 (43)	7·64 ± ·56	36·2 ± ·28 (79)	3·67 ± ·20
	XXII. Hindus from Madras Presidency	20·1 ± ·24 (49)	2·06 ± ·14	45·7 ± 1·04 (46)	10·44 ± ·73	39·6 ± 1·47* (10)	—	36·8 ± ·34 (48)	3·44 ± ·24
	XXIII (*a*). Bengalese	20·5 ± ·15 (115)	2·44 ± ·11	46·3 ± ·82 (114)	13·03 ± ·58	39·3 ± ·59 (52)	6·25 ± ·41	35·4 ± ·24 (107)	3·74 ± ·17
	XXIII (*b*). Hindus: Bihar and Orissa	20·2 ± ·20 (37)	1·78 ± ·14	47·9 ± 1·14 (32)	9·55 ± ·81	41·5 ± 1·10* (18)	—	35·9 ± ·41 (32)	3·43 ± ·29
	XXIII (*c*). Hindus: United Provinces	19·8 ± ·32* (27)	—	47·1 ± 1·59* (27)	—	41·9 ± 1·65* (8)	—	35·0 ± ·50* (24)	—
Differences of Means	XXI—XXII	+ ·9 ± ·31		+3·7 ± 1·38		+3·9 ± 1·67		− ·6 ± ·44	
	XXI—XXIII (*a*)	+ ·5 ± ·25		+3·1 ± 1·23		+4·2 ± ·98		+ ·8 ± ·37	
	XXI—XXIII (*b*)	+ ·8 ± ·28		+1·5 ± 1·46		+2·0 ± 1·35		+ ·3 ± ·50	
	XXI—XXIII (*c*)	+1·2 ± ·38		+2·3 ± 1·83		+1·6 ± 1·82		+1·2 ± ·57	
	XXII—XXIII (*a*)	− ·4 ± ·29		− ·6 ± 1·32		+ ·3 ± 1·58		+1·4 ± ·41	
	XXII—XXIII (*b*)	− ·1 ± ·31		−2·2 ± 1·54		−1·9 ± 1·84		+ ·9 ± ·53	
	XXII—XXIII (*c*)	+ ·3 ± ·40		−1·4 ± 1·90		−2·3 ± 2·21		+1·8 ± ·60	
	XXIII (*a*)—XXIII (*b*)	+ ·3 ± ·25		−1·6 ± 1·40		−2·2 ± 1·24		− ·5 ± ·48	
	XXIII (*a*)—XXIII (*c*)	+ ·7 ± ·35		− ·8 ± 1·79		−2·6 ± 1·75		+ ·4 ± ·56	
	XXIII (*b*)—XXIII (*c*)	+ ·4 ± ·37		+ ·8 ± 1·96		− ·4 ± 1·98		+ ·9 ± ·65	

* The probable error of a mean index based on less than 30 skulls was found by using an average σ found from the weighted values of the squared standard deviations $\left(\bar{\sigma}=\sqrt{\frac{n_1\sigma_1^2+n_2\sigma_2^2+\dots}{n_1+n_2+\dots}}\right)$ for all the series in the table having the same index based on more than 30 skulls.

(4) *Sexual Differences of Mean Values.* The material dealt with consists of 83 male and 48 female series, and in general the female series are considerably shorter than the corresponding male series. Table I gives their means for the eight absolute measurements and four indices together with the probable errors. The sexual differences of these constants may be considered first. The sex ratios (i.e. the male means divided by the female means) are given in Table II for the eight absolute measurements in all cases where the means for both sexes are based on 20 or more skulls. Values have been given previously for a number of the larger calvarial arcs and chords in the case of seven European and African series*. These show that the ratio may be peculiar to the particular character considered, but the range found for all of them only extended from 1·024 to 1·066. It can be seen from Table II that all the sex ratios for the internal biorbital breadth (*IOW*)—a calvarial chord—fall within the same range. Little significance can be attached to such a range, of course, since the extreme values may be affected appreciably by errors of random sampling, or the inclusion of any series for which the male and female means do not represent precisely the same racial type is likely to extend the extremes considerably. While showing a few ratios less than 1·024 or greater than 1·066, most of these for *MOW*, *GB*, Sub. *IOW* and Sub. *GB* lie between those values. As far as can be seen the sex ratios for these characters are not peculiar, and the subtenses to *IOW* and *GB* have approximately the same values as the chords. But for the remaining measurements the position is different. For the simotic chord (*SC*) there are seven ratios less than unity, the female mean being greater than the male, and 14 of the 27 ratios are less than 1·024. The simotic subtense (*SS*), however, has only one value less than 1·024, but 24 greater than 1·066. For every series the ratio for the subtense is greater than that for the chord. The last character, the rhinal subtense (Sub. *MOW*), also has high sex ratios, the lowest of the 11 that can be given being 1·078, while again the value for the subtense is invariably greater than the value for the chord (*MOW*) in the case of the same series. We see then that the simotic chord shows peculiarly small sex ratios while the simotic and rhinal subtenses are also distinguished, but on account of their peculiarly large ratios. The interracial averages shown in Table II bear out these conclusions. The breadth of the nasal bones shows relatively small differences between man and woman, but there are relatively large sexual differences between measurements of the projection of the nasal bones, both at the bridge of the nose and at their lower extremity. It may be anticipated that large sexual differences will be found between mean simotic and rhinal indices for the same series, and possibly no markedly significant sexual differences between the mean frontal and premaxillary indices. The fact that these relations do hold can be appreciated most readily from Figs. 1 and 2. The first figure shows the interracial correlation between the simotic and rhinal indices, all available means based on 15 or more skulls being shown, and it can be seen at a glance that in the case of both these indices every male mean is greater than every corresponding female value. This is a particularly striking relation since several of the series are

* See *Biometrika*, Vol. xxiii. pp. 214 and 275, 1931.

TABLE I. *Mean Measurements of Facial Flatness for different Racial Series*.*

Continent	Serial Number	Series	Sex	Absolute Measurements								Indices			
				IOW	Sub. *IOW*	*SC*	*SS*	*MOW*	Sub. *MOW*	*GB*	Sub. *GB*	Frontal	Simotic	Rhinal	Pre-maxillary
A. Europe	I	English: Long Barrow	♂	98·5 (18)	18·4	10·1 (13)	5·3	—	—	94·5 (9)	33·6	18·9 ± ·34	52·8	—	35·6
	II	Romano-British ...	„	98·5 (86)	18·3	9·3 (82)	4·6	53·6 (30)	23·3	94·8 (64)	32·8	18·5 ± ·16	50·5 ± ·93	43·8 ± ·69	35·3 ± ·25
	III	Anglo-Saxon	„	98·1 (92)	18·2	9·1 (77)	4·8	52·0 (22)	23·4	95·2 (76)	33·4	18·5 ± ·15	52·4 ± ·71	45·3 ± ·83	35·1 ± ·28
	IV	English: Mediaeval ...	„	98·2 (43)	17·7	9·4 (36)	4·5	54·8 (19)	21·7	95·2 (35)	32·6	17·9 ± ·19	47·2 ± 1·44	39·9 ± ·89	34·3 ± ·42
	V	English: Farringdon St.	„	98·2 (75)	18·6	9·2 (81)	4·6	54·1 (37)	22·9	92·4 (53)	32·7	19·0 ± ·17	50·7 ± ·96	42·7 ± ·68	35·6 ± ·34
	VI	English: Spitalfields ...	„	98·1 (248)	18·0	9·7 (126)	4·5	54·2 (50)	22·0	93·4 (73)	32·2	18·4 ± ·07	47·4 ± ·71	40·7 ± ·49	34·6 ± ·24
	VII	Irish	„	99·0 (65)	19·2	8·6 (65)	4·6	54·2 (44)	24·0	93·8 (57)	34·2	19·3 ± ·20	53·1 ± ·98	44·6 ± ·54	36·4 ± ·24
	VIII	Scottish	„	99·1 (35)	18·7	9·1 (32)	4·8	54·4 (16)	24·3	93·2 (31)	33·9	18·8 ± ·26	52·8 ± 1·12	44·9 ± ·97	36·4 ± ·38
	IX	British (Soldiers) ...	„	97·8 (113)	18·5	8·9 (111)	4·7	54·3 (84)	24·1	93·1 (101)	33·6	19·0 ± ·16	52·7 ± ·84	44·4 ± ·45	36·1 ± ·21
	X	North Germans ...	„	99·5 (32)	18·7	9·6 (30)	4·9	52·9 (13)	25·6	94·7 (31)	33·4	18·8 ± ·28	51·5 ± 1·47	48·6	35·4 ± ·44
	XI	French	„	96·6 (46)	18·7	9·5 (43)	4·6	53·8 (11)	24·8	91·7 (39)	32·4	19·3 ± ·21	49·2 ± ·92	46·4	35·4 ± ·44
	XII	Italian	„	98·1 (100)	19·3	9·7 (100)	4·5	53·6 (52)	24·2	94·4 (94)	32·9	19·7 ± ·16	47·2 ± ·77	45·4 ± ·54	34·8 ± ·24
	XIII	Norwegian	„	97·5 (19)	18·9	9·2 (18)	4·9	53·9 (2)	24·6	92·2 (19)	33·3	19·4 ± ·33	53·0 ± 1·85	45·7	36·3 ± ·51
	XIV	Swede	„	98·7 (30)	19·3	8·8 (29)	4·5	55·7 (7)	25·4	93·3 (29)	33·0	19·6 ± ·29	52·2 ± 1·46	45·6	35·7 ± ·41
	XV	Finn	„	100·0 (23)	19·0	10·3 (22)	4·8	52·8 (6)	20·9	97·2 (23)	33·3	19·0 ± ·30	47·2 ± 1·67	39·7	34·3 ± ·47
	XVI	Dutch	„	99·4 (51)	19·1	9·0 (51)	4·7	53·6 (22)	25·7	92·8 (45)	33·2	19·2 ± ·24	53·0 ± 1·05	48·3 ± ·83	35·7 ± ·33
	XVII	Ancient Greek	„	98·5 (12)	19·4	10·5 (11)	5·3	—	—	97·6 (9)	33·4	19·7	51·4	—	34·4
	XVIII	Modern Greek	„	97·3 (33)	19·0	9·4 (32)	4·5	56·2 (11)	24·6	95·0 (25)	33·1	19·5 ± ·27	47·8 ± 1·20	44·3	34·9 ± ·45
	XIX	Czech	„	97·7 (23)	17·9	10·0 (23)	4·6	54·3 (11)	23·0	93·0 (19)	33·6	18·3 ± ·30	46·0 ± 1·64	42·5	36·4 ± ·51
	XX	Russian	„	98·4 (35)	18·4	9·1 (36)	4·5	53·2 (16)	23·8	95·5 (33)	34·3	18·7 ± ·23	50·0 ± 1·32	45·2 ± ·97	35·9 ± ·31
	II	Romano-British ...	♀	95·1 (45)	17·7	9·0 (42)	4·2	51·7 (16)	20·6	89·5 (39)	31·4	18·6 ± ·22	45·1 ± 1·12	40·2 ± ·89	35·0 ± ·38
	III	Anglo-Saxon	„	94·0 (45)	17·2	8·7 (41)	4·0	51·9 (11)	21·3	89·9 (40)	32·0	18·2 ± ·16	45·9 ± ·98	41·3	35·6 ± ·30
	IV	English: Mediaeval ...	„	94·2 (11)	16·4	8·3 (10)	3·5	—	—	90·6 (9)	30·1	17·4	40·0	—	33·4
	V	English: Farringdon St.	„	94·2 (75)	17·6	9·1 (76)	4·0	52·1 (47)	20·5	86·7 (51)	30·7	18·7 ± ·17	44·6 ± ·79	39·4 ± ·52	35·6 ± ·38
	VI	English: Spitalfields ...	„	93·5 (94)	17·2	9·6 (31)	4·3	53·7 (14)	20·1	87·9 (19)	30·8	18·2 ± ·12	45·3 ± 1·41	37·8	34·8 ± ·55
	XI	French	„	94·3 (19)	18·8	10·6 (18)	4·6	51·6 (4)	21·1	87·7 (15)	31·5	19·9 ± ·31	43·3 ± 1·66	40·7	35·9 ± ·62
	XII	Italian	„	92·2 (32)	18·3	9·4 (31)	3·9	50·7 (20)	20·9	87·8 (24)	31·1	19·8 ± ·32	41·7 ± 1·29	41·2 ± ·80	35·4 ± ·48
	XVI	Dutch	„	94·4 (10)	18·3	8·5 (10)	4·2	48·2 (4)	22·8	88·2 (8)	32·2	19·3	50·2	47·4	36·5
	XIX	Czech	„	92·4 (10)	17·5	8·9 (8)	4·0	54·6 (3)	21·1	86·1 (9)	30·7	18·9	44·7	38·7	35·7

B. Africa	I	Guanche	♂	89·1 (25)	18·8	10·2 (24)	4·8	59·6 (5)	24·9	98·4 (23)	32·1	19·0± ·30	47·4±1·28	41·8	32·6± ·50
	II	Nubian	„	94·4 (69)	17·5	11·7 (55)	4·2	54·1 (8)	18·6	95·6 (40)	33·1	18·3± ·20	36·3± ·74	34·6	34·5± ·39
	III	Egyptian: Badari ...	„	94·1 (32)	16·6	11·1 (34)	4·2	54·4 (20)	18·8	94·6 (34)	35·1	17·6± ·25	38·7± ·82	34·8± ·63	37·2± ·39
	IV	Egyptian: Sedment ...	„	95·5 (40)	17·5	9·5 (37)	3·9	54·1 (21)	21·8	93·9 (37)	36·2	18·3± ·25	41·1±1·11	40·3± ·62	38·5± ·38
	V	Egyptian: Kerma ...	„	96·2 (117)	17·5	10·4 (99)	3·8	58·8 (22)	19·3	94·9 (103)	33·9	18·2± ·15	36·5± ·62	32·9± ·60	35·8± ·24
	VI	Negro: Nigeria ...	„	99·3 (52)	17·5	9·4 (49)	2·7	59·0 (23)	17·5	95·0 (46)	34·6	17·6± ·17	29·5± ·93	29·8± ·59	36·4± ·36
	VII	Gaboon Negro	„	99·3 (76)	17·4	9·3 (71)	2·8	62·1 (49)	17·9	95·7 (63)	36·9	17·6± ·17	29·9± ·70	28·9± ·40	38·6± ·28
	VIII	Angoni	„	99·6 (17)	17·1	8·7 (22)	2·5	63·1 (12)	17·8	95·2 (14)	37·1	17·1± ·37	30·1±1·34	28·4	39·1
	IX	Ashanti	„	99·4 (33)	17·3	8·8 (34)	2·9	61·1 (18)	18·3	95·3 (30)	36·5	17·4± ·21	34·0±1·23	30·3± ·67	38·4± ·39
	X	Teita Negro	„	99·1 (55)	17·2	9·4 (39)	2·2	63·5 (27)	17·1	98·5 (31)	33·1	17·3± ·20	22·4± ·83	26·9± ·54	33·7± ·33
	XI	Zulu	„	102·4 (17)	18·6	9·4 (17)	2·6	60·1 (6)	18·4	97·6 (16)	37·1	18·1± ·37	28·3±1·52	30·8	38·1± ·59
	XII	Congo Negro	„	98·6 (60)	17·1	9·5 (60)	2·5	59·8 (17)	17·8	93·9 (42)	33·3	17·3± ·19	26·5± ·72	29·9± ·68	35·5± ·44
	XIII	Bushman	„	95·6 (21)	14·8	6·8 (21)	1·4	58·4 (10)	11·5	92·7 (16)	31·5	15·5± ·33	20·7±1·37	19·7	34·1± ·59
	XIV	Hottentot	„	97·3 (9)	16·7	7·9 (9)	2·3	60·1 (4)	13·5	93·7 (8)	32·9	17·0	29·3	22·7	35·1
	XV	Kaffir	„	103·0 (54)	17·8	8·5 (54)	2·8	60·9 (25)	18·3	96·5 (51)	35·4	17·3± ·22	33·6±1·10	30·4± ·56	36·8± ·34
	I	Guanche	♀	93·5 (26)	17·5	10·3 (27)	4·0	53·5 (6)	21·1	93·0 (22)	30·0	18·7± ·27	38·7±1·11	40·0	32·2± ·45
	II	Nubian	„	92·0 (61)	17·0	10·5 (38)	3·2	58·7 (4)	19·7	91·9 (35)	32·5	18·4± ·16	30·2± ·67	33·6	35·3± ·29
	III	Egyptian: Badari ...	„	90·3 (19)	16·0	10·6 (18)	4·0	53·6 (9)	18·0	90·9 (15)	33·4	17·7± ·32	37·5±1·36	33·8	36·7± ·54
	IV	Egyptian: Sedment ...	„	90·0 (30)	15·7	10·0 (29)	3·8	51·5 (21)	19·4	89·4 (28)	34·0	17·4± ·25	37·9±1·08	37·7± ·60	37·9± ·40
	V	Egyptian: Kerma ...	„	92·3 (94)	16·3	9·9 (74)	3·2	54·1 (11)	16·3	90·7 (83)	33·1	17·7± ·15	32·1± ·74	30·3	36·5± ·22
	VI	Negress: Nigeria ...	„	95·9 (19)	16·9	9·7 (17)	2·1	61·3 (6)	17·8	91·2 (15)	33·5	17·6± ·32	22·4±1·40	29·6	36·8± ·54
	VII	Gaboon Negress ...	„	94·8 (60)	16·7	9·7 (55)	2·5	59·0 (44)	16·6	90·6 (51)	35·2	17·6± ·18	26·1± ·77	28·2± ·40	38·9± ·31
	IX	Ashanti	„	94·0 (31)	16·9	8·0 (30)	2·1	58·3 (14)	16·1	88·5 (23)	35·1	18·0± ·25	26·4±1·35	27·6	39·7± ·44
	X	Teita Negress	„	94·4 (62)	16·6	10·2 (49)	1·9	61·7 (31)	14·9	94·3 (37)	32·1	17·6± ·18	18·6± ·69	24·2± ·52	34·1± ·41
	XII	Congo Negress	„	95·2 (30)	16·2	9·5 (25)	2·3	59·3 (9)	16·2	92·9 (23)	32·9	17·0± ·23	25·7±1·16	27·5	35·4± ·44
	XIII	Bushwoman	„	92·3 (17)	14·5	7·9 (16)	1·4	56·7 (9)	10·9	86·6 (15)	29·7	15·7± ·34	16·9±1·45	19·2	34·6± ·54
	XIV	Hottentot	„	94·5 (8)	15·7	7·6 (8)	2·0	53·8 (2)	15·0	88·8 (8)	31·7	16·6	27·5	27·9	35·8
	XV	Kaffir	„	99·1 (10)	17·1	7·7 (9)	1·8	56·1 (5)	15·8	91·6 (10)	33·2	17·3	24·3	28·3	36·4
C. America	I	Eskimo	♂	99·1 (51)	14·0	5·7 (50)	2·4	51·9 (20)	17·1	103·3 (45)	32·1	14·1± ·19	42·5± ·85	33·3± ·73	31·4± ·33
	II	Peruvian	„	97·3 (171)	16·7	9·2 (166)	3·8	57·8 (43)	22·7	99·8 (143)	34·2	17·2± ·09	41·9± ·50	39·4± ·50	34·4± ·17
	III	North American Indian	„	100·2 (45)	16·8	8·5 (46)	2·7	55·7 (20)	20·3	101·0 (41)	33·4	16·7± ·18	31·4±1·13	36·5± ·73	33·5± ·31
	IV	Patagonian	„	102·0 (20)	18·0	8·4 (20)	4·2	65·6 (2)	23·6	102·8 (19)	37·0	17·6± ·28	49·0±1·47	36·2	36·0± ·47
	V	Fuegian	„	99·0 (12)	16·5	7·9 (12)	3·2	58·0 (5)	22·9	99·3 (12)	35·4	16·7	42·8	39·4	35·6
	I	Eskimo	♀	94·3 (30)	13·6	5·4 (29)	1·9	48·3 (8)	15·1	95·8 (27)	31·6	14·4± ·21	35·7± ·92	31·1	33·0± ·38
	II	Peruvian	„	93·4 (114)	15·5	9·0 (107)	3·2	53·6 (20)	20·3	94·4 (97)	33·3	16·7± ·13	36·2± ·48	38·1± ·79	35·3± ·20
	III	North American Indian	„	95·8 (26)	16·5	8·3 (26)	2·4	54·8 (13)	17·7	95·3 (24)	32·6	17·2± ·27	29·5± ·97	32·4	34·3± ·41

* The serial numbers in the second column of this table are those used in the description of the material in section (3) of the text. The numbers in round brackets are the numbers of skulls on which the means are based, and for every series the chord and the corresponding subtense and index are for the same skulls, so the number is only given in the case of the chord. The probable errors of mean indices based on 30 or more skulls were found by using the standard deviations for the same series; for those based on 15—29 skulls the standard deviations used are the weighted average values for all series belonging to the same continental area, and where the means are based on fewer than 15 skulls the probable errors are not given. All the standard deviations used are given in Table V.

TABLE I—(*continued*).

Continent	Serial Number	Series	Sex	Absolute Measurements								Indices			
				IOW	Sub. *IOW*	*SC*	*SS*	*MOW*	Sub. *MOW*	*GB*	Sub. *GB*	Frontal	Simotic	Rhinal	Pre-maxillary
D. Oceania	I	Maori	♂	99·1 (89)	17·0	7·2 (87)	3·1	53·8 (38)	21·1	98·8 (76)	35·4	17·3 ± ·14	43·2 ± ·88	39·4 ± ·56	36·3 ± ·27
	II	Moriori	,,	98·9 (43)	17·3	6·5 (42)	2·8	57·8 (16)	22·1	102·6 (39)	37·0	17·4 ± ·19	43·4 ± 1·24	38·6 ± ·81	36·2 ± ·31
	III	New British	,,	99·6 (39)	16·1	8·0 (39)	2·7	59·4 (33)	16·5	96·3 (35)	37·3	16·1 ± ·20	33·4 ± 1·14	27·9 ± ·41	38·9 ± ·27
	IV	Marquesas Islands	,,	98·5 (22)	17·6	7·4 (22)	2·9	57·1 (5)	18·7	99·4 (22)	34·5	17·9 ± ·33	40·6 ± 1·61	33·6	34·7 ± ·48
	V	Easter Island	,,	98·1 (41)	17·7	7·8 (44)	3·1	59·5 (21)	21·2	98·0 (36)	35·7	18·0 ± ·20	40·0 ± 1·25	36·0 ± ·71	36·2 ± ·38
	VI	Kanaka: Sandwich Is.	,,	97·7 (67)	16·0	7·4 (63)	2·7	58·0 (47)	19·1	98·2 (54)	33·5	16·4 ± ·17	36·7 ± 1·00	33·8 ± ·53	34·3 ± ·26
	VII	New Caledonian	,,	98·1 (15)	16·9	8·5 (15)	3·0	60·0 (4)	21·0	97·1 (14)	34·6	17·2 ± ·39	35·4 ± 1·95	34·9	35·7
	VIII	Solomon Islands	,,	98·8 (27)	17·4	8·0 (26)	3·0	59·0 (10)	18·5	97·0 (23)	34·9	17·6 ± ·29	37·6 ± 1·48	31·3	36·0 ± ·47
	IX	New Hebrides Islands	,,	99·2 (36)	17·7	8·6 (34)	3·1	57·9 (15)	20·0	96·6 (33)	35·7	17·8 ± ·25	36·1 ± 1·17	34·9 ± ·84	37·0 ± ·44
	X	Fijian	,,	99·8 (30)	18·4	8·6 (30)	3·4	57·6 (14)	21·1	97·3 (26)	36·3	18·4 ± ·24	39·8 ± 1·34	36·8	37·3 ± ·44
	XI	Papuan	,,	97·3 (76)	16·7	8·1 (78)	3·2	57·8 (47)	18·7	96·7 (65)	34·6	17·2 ± ·14	39·6 ± ·77	32·6 ± ·50	35·9 ± ·30
	XII	Cook Group Islands	,,	97·8 (17)	17·3	8·2 (15)	3·3	59·1 (4)	20·5	95·9 (19)	34·6	17·6 ± ·37	40·1 ± 1·95	34·6	36·2 ± ·52
	XIII	Tasmanian	,,	99·5 (21)	18·6	8·4 (20)	3·4	—	—	91·4 (17)	38·7	18·7 ± ·33	41·7 ± 1·69	—	42·3 ± ·55
	XIV	Australian: Northern Territory	,,	98·5 (18)	18·5	8·6 (16)	3·5	54·2 (12)	20·1	89·8 (17)	37·2	18·8 ± ·36	41·9 ± 1·89	37·2	41·5 ± ·55
	XV	Australian: Queensland	,,	101·8 (36)	18·8	9·1 (36)	3·4	60·0 (18)	18·1	94·3 (35)	37·2	18·5 ± ·28	37·5 ± 1·01	30·1 ± ·67	39·5 ± ·39
	XVI	Australian: "Other" regions ...	,,	101·5 (121)	18·7	9·3 (114)	4·1	57·7 (50)	19·6	92·9 (103)	37·0	18·4 ± ·18	44·4 ± ·63	34·0 ± ·44	39·8 ± ·24
	I	Maori	♀	93·8 (23)	15·4	7·0 (25)	2·5	54·2 (11)	17·9	93·5 (20)	34·5	16·4 ± ·28	36·8 ± 1·26	33·2	37·0 ± ·52
	II	Moriori	,,	94·3 (26)	17·0	7·6 (25)	2·6	56·3 (7)	20·5	95·9 (23)	34·9	18·0 ± ·26	34·6 ± 1·26	36·9	36·4 ± ·48
	III	New British	,,	94·0 (24)	15·9	7·9 (24)	2·1	55·4 (18)	14·4	91·1 (22)	35·4	16·9 ± ·27	26·1 ± 1·28	26·1 ± ·63	39·0 ± ·49
	IV	Marquesas Islands	,,	89·9 (16)	15·5	7·2 (16)	2·2	55·1 (10)	17·3	92·3 (14)	33·7	17·1 ± ·33	29·9 ± 1·57	31·5	36·5
	V	Easter Island	,,	93·3 (18)	16·2	8·1 (21)	2·5	56·4 (10)	18·6	91·5 (8)	32·6	17·3 ± ·31	30·6 ± 1·37	33·0	35·7
	VI	Kanaka: Sandwich Is.	,,	92·8 (60)	15·0	7·2 (60)	2·1	54·4 (37)	16·7	91·8 (57)	32·7	16·0 ± ·18	28·6 ± ·96	30·7 ± ·42	35·7 ± ·27
	X	Fijian	,,	93·3 (13)	16·5	7·4 (11)	2·8	53·0 (4)	19·9	91·0 (13)	34·5	17·6	37·3	37·4	37·9
	XI	Papuan	,,	93·3 (49)	16·5	8·2 (49)	2·8	54·9 (34)	16·9	91·6 (40)	33·7	17·6 ± ·18	34·0 ± ·81	31·1 ± ·46	36·9 ± ·43
	XIII	Tasmanian	,,	95·6 (12)	16·9	8·0 (11)	2·1	—	—	89·4 (11)	34·9	17·6	25·1	—	38·7
	XV	Australian: Queensland ...	,,	96·1 (12)	17·5	8·5 (12)	3·1	57·5 (5)	17·2	88·8 (11)	35·3	18·2	35·5	29·9	39·9
	XVI	Australian: "Other" regions ...	,,	95·9 (80)	17·3	8·8 (74)	3·0	55·7 (27)	16·8	88·7 (71)	35·5	18·0 ± ·15	34·5 ± ·65	30·5 ± ·52	40·0 ± ·27

E. Asia	I	Chinese	♂	96·1 (136)	15·1	7·6 (137)	2·4	54·6 (109)	17·0	99·3 (121)	34·0	15·8±·13	31·7± ·62	31·5±·34	34·2±·22
	II	Burmese	„	98·4 (84)	15·7	8·9 (82)	3·0	57·6 (57)	16·9	101·5 (70)	33·7	16·0±·17	33·1± ·67	29·4±·34	33·2±·24
	III	Japanese	„	98·4 (17)	15·9	7·9 (17)	2·4	55·1 (12)	17·4	99·7 (16)	35·5	16·2±·36	31·7±1·73	31·9	35·6±·60
	IV	Aino	„	98·5 (8)	17·0	8·6 (8)	3·2	54·1 (4)	17·4	100·9 (7)	34·3	17·2	36·6	32·2	34·0
	V	Tibetan *A*	„	96·5 (37)	15·5	8·0 (35)	2·5	56·0 (34)	17·3	98·6 (36)	31·5	16·1±·23	31·8± ·95	31·0±·66	32·0±·36
	VI	Tibetan *B*	„	99·5 (15)	16·2	7·7 (15)	2·6	56·4 (6)	17·7	100·4 (14)	35·4	16·2±·38	34·6±1·85	32·0	35·4
	VII	Nepalese	„	94·3 (46)	15·9	8·1 (47)	3·0	56·3 (39)	18·0	97·4 (44)	32·2	16·9±·26	37·7±1·43	31·5±·63	33·1±·30
	VIII	Malay	„	96·0 (10)	15·0	7·8 (10)	2·7	53·9 (7)	17·0	96·6 (10)	33·7	15·6	33·9	31·5	35·1
	IX	Andamanese	„	91·6 (31)	15·8	9·4 (25)	2·4	52·6 (8)	18·0	94·8 (27)	32·8	17·2±·22	25·3±1·43	34·2	34·6±·46
	X	Nicobarese	„	94·7 (14)	15·3	9·2 (14)	2·6	55·5 (11)	14·8	95·7 (10)	32·3	16·2	27·0	26·6	33·8
	XI	Naga	„	96·9 (20)	15·8	8·5 (16)	2·9	—	—	102·2 (10)	34·4	16·2±·31	34·9±1·79	—	33·8
	XII	Javanese	„	97·0 (176)	15·4	8·7 (172)	2·8	56·7 (148)	17·3	100·5 (166)	34·6	15·8±·11	32·9± ·50	30·5±·27	34·5±·20
	XIII	Sumatran	„	96·4 (56)	15·9	8·4 (56)	2·7	55·6 (52)	18·0	99·5 (50)	35·5	16·5±·14	32·6± ·91	32·8±·55	35·8±·34
	XIV	Celebes (Natives)	„	99·0 (52)	16·3	8·7 (52)	2·7	55·6 (45)	17·6	100·6 (49)	34·3	16·4±·21	31·7± ·70	31·7±·43	34·2±·33
	XV	Sarawak (Natives)	„	96·4 (55)	16·2	8·0 (55)	2·6	55·7 (25)	17·3	98·7 (43)	32·5	16·8±·19	33·7± ·91	31·3±·70	32·9±·47
	XVI	Dayak	„	96·5 (55)	15·9	8·5 (50)	2·7	56·6 (44)	17·6	101·2 (46)	34·2	16·5±·18	31·8± ·94	31·3±·45	33·8±·33
	XVII	Tagal	„	97·6 (31)	16·5	9·6 (27)	3·4	59·2 (11)	17·0	100·1 (23)	34·1	16·9±·20	36·0±1·38	28·9	34·1±·50
	XVIII	Aëta	„	96·4 (33)	15·9	8·7 (26)	3·1	58·3 (19)	16·4	98·8 (27)	34·2	16·5±·19	35·3±1·40	28·1±·81	34·7±·46
	XIX	Singalese	„	96·1 (34)	20·1	8·8 (34)	4·0	56·8 (28)	22·8	94·0 (26)	36·0	21·0±·26	45·8±1·20	40·3±·66	38·4±·47
	XX	Veddah	„	93·7 (21)	19·3	8·5 (20)	3·8	57·0 (12)	20·8	91·7 (13)	33·8	20·6±·32	44·8±1·60	36·6	37·6
	XXI	Punjabi	„	95·8 (93)	20·2	9·1 (92)	4·5	55·6 (43)	24·1	93·8 (79)	33·7	21·0±·20	49·4± ·91	43·5±·79	36·2±·28
	XXII	Hindu: Madras	„	94·7 (49)	19·0	8·7 (46)	3·9	55·8 (10)	22·0	93·0 (48)	34·2	20·1±·24	45·7±1·04	39·6	36·8±·34
	XXIII	Hindu: N.E. India	„	95·4 (179)	19·4	8·7 (173)	4·1	57·1 (78)	22·7	94·5 (163)	33·4	20·3±·12	46·7± ·57	39·9±·38	35·4±·18
	XXIV	Dravidian	„	93·2 (36)	18·9	8·7 (35)	3·7	54·8 (2)	21·6	93·0 (31)	34·1	20·3±·24	43·1±1·19	39·5	36·6±·40
	XXV	Afghan	„	97·6 (18)	20·0	9·8 (17)	5·0	55·3 (8)	24·0	96·9 (16)	35·5	20·5±·35	52·0±1·73	43·4	36·7±·60
	XXVI	Arab	„	97·5 (9)	18·9	9·4 (9)	4·2	54·1 (5)	22·8	92·3 (9)	34·3	19·4	45·5	42·5	37·2
	XXVII	Syrian and Palestinian (Natives)	„	95·5 (25)	18·7	10·0 (20)	5·2	56·7 (4)	27·3	92·4 (16)	33·3	19·5±·30	52·4±1·60	48·4	36·1±·60
	I	Chinese	♀	91·3 (17)	14·6	7·0 (16)	1·8	52·1 (12)	15·8	93·5 (16)	33·2	16·0±·32	25·8±1·54	30·4	35·5±·53
	II	Burmese	„	93·3 (38)	14·3	8·5 (37)	2·5	55·2 (15)	14·0	97·1 (30)	31·3	15·3±·22	29·8± ·94	25·4±1·00	32·5±·34
	IX	Andamanese	„	89·1 (25)	15·3	9·2 (24)	2·2	51·8 (12)	16·8	89·7 (25)	31·8	17·1±·26	23·9±1·26	32·4	35·4±·42
	XI	Naga	„	92·2 (15)	14·7	8·2 (14)	2·2	55·8 (3)	14·0	94·6 (10)	32·0	15·9±·34	25·5	25·6	33·9
	XII	Javanese	„	93·3 (34)	14·3	8·7 (34)	2·4	55·6 (30)	15·7	95·5 (30)	33·6	15·4±·20	27·7± ·89	28·5±·58	35·2±·45
	XV	Sarawak (Natives)	„	91·9 (24)	15·7	8·6 (23)	2·6	51·8 (9)	13·8	92·5 (16)	31·4	16·9±·27	31·7±1·28	26·7	33·9±·53
	XVI	Dayak	„	93·0 (18)	15·8	8·0 (17)	2·3	53·9 (9)	16·1	95·4 (12)	33·1	17·0±·31	28·4±1·49	29·8	34·6
	XVII	Tagal	„	91·2 (19)	14·9	8·4 (17)	2·4	50·5 (8)	15·8	93·8 (17)	32·2	16·3±·30	28·1±1·49	31·6	34·3±·51
	XVIII	Aëta	„	92·8 (14)	14·7	8·0 (13)	2·3	56·3 (6)	14·8	93·4 (11)	31·7	15·8	29·7	26·9	34·1
	XIX	Singalese	„	92·8 (13)	19·0	8·6 (13)	3·3	54·9 (11)	19·4	92·2 (11)	34·7	20·6	37·5	35·4	37·6
	XX	Veddah	„	88·7 (14)	17·5	8·3 (12)	3·1	54·2 (6)	20·8	86·5 (10)	32·1	19·6	38·2	38·5	37·0
	XXIII	Hindu: N.E. India	„	91·8 (50)	18·1	8·5 (48)	3·5	53·3 (25)	18·8	92·2 (46)	31·8	19·7±·20	42·5±1·02	35·7±·78	35·3±·29

TABLE II.

Sex Ratios of the Absolute Measurements (all means used based on Twenty or more Skulls).

Continental Area	Serial No.	Series*	Subtenses				Chords			
			Sub. *IOW*	*SS*	Sub. *MOW*	Sub. *GB*	*IOW*	*SC*	*MOW*	*GB*
A. European Series	III	Anglo-Saxon	1·058	1·200	—	1·044	1·044	1·046	—	1·059
	V	English: Farringdon St.	1·057	1·150	1·117	1·065	1·042	1·011	1·039	1·065
	XII	Italian	1·055	1·154	1·158	1·058	1·064	1·032	1·057	1·075
	VI	English: Spitalfields ...	1·047	1·047	—	—	1·049	1·010	—	—
	II	Romano-British	1·034	1·095	—	1·045	1·026	1·033	—	1·059
		Average ratios	1·050	1·129	1·137	1·053	1·047	1·026	1·048	1·065
B. African Series	IV	Egyptian: Sedment ...	1·115	1·026	1·124	1·065	1·061	·950	1·050	1·050
	I	Guanche	1·074	1·190	—	1·070	1·060	·990	—	1·058
	V	Egyptian: Kerma ...	1·074	1·188	—	1·024	1·042	1·051	—	1·046
	XII	Congo Negro	1·056	1·087	—	1·012	1·036	1·000	—	1·011
	VII	Gaboon Negro	1·042	1·104	1·078	1·048	1·047	·959	1·053	1·057
	X	Negro: Teita Hills ...	1·036	1·158	1·148	1·031	1·050	·922	1·029	1·045
	II	Nubian	1·029	1·313	—	1·018	1·026	1·114	—	1·040
	IX	Ashanti	1·024	1·381	—	1·040	1·057	1·100	—	1·077
		Average ratios	1·056	1·181	1·117	1·039	1·047	1·011	1·044	1·048
C. American Series	II	Peruvian	1·077	1·188	1·118	1·027	1·042	1·022	1·078	1·057
	I	Eskimo	1·029	1·263	—	1·016	1·051	1·056	—	1·078
	III	North American Indian	1·018	1·125	—	1·025	1·046	1·024	—	1·060
		Average ratios	1·041	1·192	1·118	1·023	1·046	1·034	1·078	1·065
D. Oceanic Series	I	Maori	1·104	1·240	—	1·026	1·056	1·029	—	1·057
	XI	Papuan	1·112	1·143	1·107	1·027	1·044	·988	1·053	1·056
	XVI	Australian: "Other regions"	1·081	1·367	1·167	1·042	1·058	1·057	1·036	1·047
	VI	Kanaka	1·067	1·286	1·144	1·024	1·053	1·028	1·066	1·070
	II	Moriori	1·018	1·077	—	1·060	1·049	·855	—	1·070
	III	New British	1·013	1·286	—	1·054	1·060	1·012	—	1·057
		Average ratios	1·049	1·233	1·139	1·039	1·053	·955	1·052	1·059
E. Asiatic Series	II	Burmese	1·098	1·200	—	1·077	1·055	1·047	—	1·045
	XII	Javanese	1·077	1·167	1·102	1·030	1·040	1·000	1·020	1·052
	XXIII	Hindu: N.E. India ...	1·072	1·171	1·207	1·050	1·039	1·024	1·071	1·025
	IX	Andamanese	1·033	1·091	—	1·031	1·028	1·022	—	1·057
	XV	Sarawak (Natives) ...	1·032	·963	—	—	1·049	·930	—	—
		Average ratios	1·062	1·126	1·155	1·047	1·042	1·005	1·045	1·045
		Average of all Series ...	1·053 (27)	1·173 (27)	1·134 (11)	1·050 (25)	1·048 (27)	1·012 (27)	1·040 (11)	1·055 (25)

* The series for each continental area are arranged in order according to the size of the sex ratios for the first character.

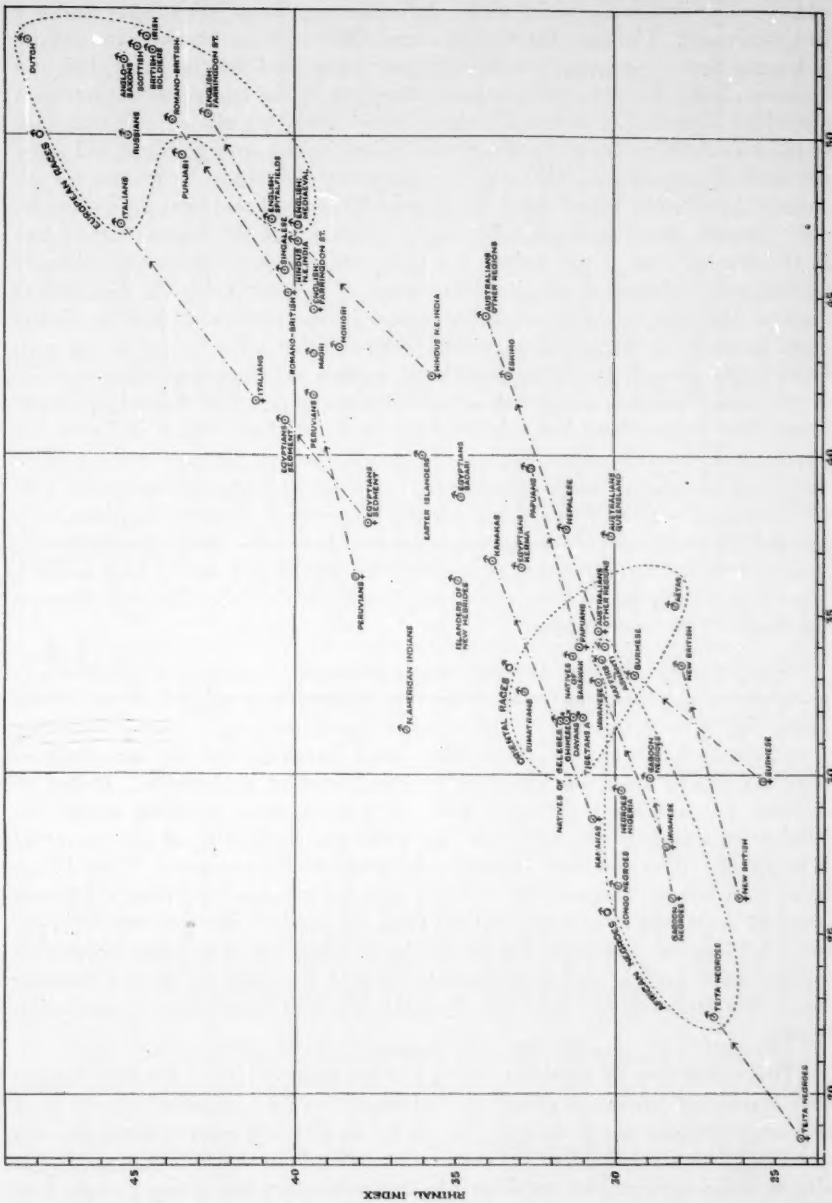

THE INTERRACIAL CORRELATION OF RHINAL AND SIMOTIC INDICES.
All means used are based on 15 or more skulls.
RHINAL INDEX
SIMOTIC INDEX
45
40
35
30
25
20
50
45
40
35
30
25
DUTCH
ENGLISH
SCOTTISH
BRITISH ISLES
IRISH
RUSSIANS
ROMANO-BRITISH
ENGLISH TRANSITION ST.
ITALIANS
PUNJABI
ENGLISH ENTHLFIELDS
SINGALESE
HINDUS N.E. INDIA
ENGLISH MEDIAEVAL
ROMANO-BRITISH
HINDUS N.E. INDIA
PARKINSON ST.
MORIORI
MAORI
ITALIANS
EGYPTIANS
PERUVIANS
HINDUS N.E. INDIA
AUSTRALIANS OTHER REGIONS
EBBING
EGYPTIANS
EARTH ISLANDERS
EGYPTIANS
PERUVIANS
NAMAQUAS
EGYPTIANS
PUNJABI
NEPALESE
AUSTRALIANS QUEENSLAND
ISLANDERS OF NEW HEBRIDES
PERUVIANS
PAPUANS
AUSTRALIANS OTHER REGIONS
AMERICAN INDIANS
ORIENTAL RACES
SUMATRANS
NATIVES OF CELEBES
BORNEO
TIBETANS
JAVANESE
BURMESE
DAYAK
NEW BRITISH
BURMESE
NATIVES OF NEW GUINEA
KAFFIRS
HEBRIDES NIGERIA
CONGO NEGROES
JAPANESE
GABOON NEGROES
NEW BRITISH
TEITA NEGROES
TEITA NEGROES

Fig. 1.

small, and it may be confidently asserted that for the same series no one of the usual cranial indices for which sexual differences are found would have shown a like uniformity. The fact that marked sexual differences are found for the simotic index was first noted by de Mérejkowsky, and it was confirmed by Ryley, Bell and Pearson. Table III gives the absolute differences of the mean indices and these quantities expressed in terms of their probable errors. It will be seen that most of the sexual differences of the simotic and rhinal indices are significant, and many are markedly significant. The degree of significance tends to be greater for the simotic than for the rhinal index, but it must be remembered that the means for the former are based on considerably larger numbers than the means for the latter in the case of most of the series. For these two characters there are evidently marked sexual differences for all modern races. Fig. 2 and Table III enable us to examine the frontal and premaxillary indices in the same way. For the former there is really no suggestion of sexual differentiation. For 15 series the male index is the greater, for 11 the position is reversed and there is absolute equality in one case. The most significant sexual difference is only 3·09 times its probable error. The premaxillary index is a measurement of shape, which indicates less similarity between the sexes. There are six series having the male mean greater, 18 having the female mean greater, and equality is shown for one series. No differences are markedly significant, but the three which exceed three times their probable errors all have the female mean greater. It is quite likely that the female index would be found invariably greater than the male in the case of large samples, but sexual differentiation here is obviously much less than that observed between the shapes of the nasal bones.

We may conclude that the male nasal bones are, on the average, decidedly more protruding than the female, while the premaxillary region tends to be less protruding in the male. The last relation will clearly not help in aiding the sexing of individual crania, and the rhinal index is of little value for the same purpose since the tips of the nasal bones are so often found to be defective. It may be doubted, too, whether the simotic index is of much value in aiding sexual discrimination owing to the fact that the intraracial variability of the character is very large. The difference between the means for the two sexes (Table III) is in all cases but one (Australians: "Other" regions) less than both male and female standard deviations for the same series (Table V), and in most cases very decidedly less. In these circumstances the simotic index can be of little value by itself in aiding "mathematical sexing," though the form of the nasal bridge is a character which may profitably be considered in conjunction with many others in anatomical sexing.

These data may be considered from another point of view. Do they suggest that sexual differences are greater for "advanced" or for "primitive" races? More and larger samples would obviously be needed to give any decisive answer to this question. The sex ratios for the simotic subtense and the sexual differences for the simotic index need only be considered in this connection, and it may be seen from

TABLE III.

Sexual Differences of the Mean Indices of Facial Flattening (all means used based on Twenty or more Skulls).

Continental Area	Serial No.	Series*	Frontal		Simotic		Rhinal		Premaxillary	
			$\Delta_{\sigma-\varphi} \pm$ p.e. Δ	$\frac{\Delta_{\sigma-\varphi}}{\text{p.e. }\Delta}$	$\Delta_{\sigma-\varphi} \pm$ p.e. Δ	$\frac{\Delta_{\sigma-\varphi}}{\text{p.e. }\Delta}$	$\Delta_{\sigma-\varphi} \pm$ p.e. Δ	$\frac{\Delta_{\sigma-\varphi}}{\text{p.e. }\Delta}$	$\Delta_{\sigma-\varphi} \pm$ p.e. Δ	$\frac{\Delta_{\sigma-\varphi}}{\text{p.e. }\Delta}$
A. European Series	III	Anglo-Saxon	+·3±·22	+1·36	+6·5±1·21	+ 5·37	—	—	−·5±·41	−1·22
	V	English: Farringdon St. ...	+·3±·24	+1·25	+6·1±1·24	+ 4·92	+3·3±·86	+3·84	0±·51	0
	II	Romano-British	−·1±·27	−·37	+5·4±1·46	+ 3·70	—	—	+·3±·46	+·65
	XII	Italian	−·1±·35	−·29	+5·5±1·50	+ 3·67	+4·2±·96	+4·37	−·6±·54	−1·11
	VI	English: Spitalfields ...	+·2±·14	+1·43	+2·1±1·58	+ 1·33	—	—	—	—
B. African Series	II	Nubian	−·1±·26	−·38	+6·1±1·00	+ 6·10	—	—	−·8±·48	−1·67
	I	Guanche	+·3±·41	+·73	+8·7±1·69	+ 5·15	—	—	+·4±·67	+·60
	V	Egyptian: Kerma ...	+·5±·21	+2·38	+4·4±·97	+ 4·54	—	—	−·7±·33	−2·12
	IX	Ashanti	−·6±·33	−1·82	+7·6±1·83	+ 4·15	—	—	−1·3±·59	−2·20
	VII	Gaboon Negro	0±·24	0	+3·8±1·03	+ 3·69	+·7±·45	+1·56	−·3±·41	−·73
	X	Negro: Teita Hills ...	−·3±·27	−1·11	+3·8±1·08	+ 3·52	+2·7±·75	+3·60	−·4±·55	−·73
	IV	Egyptian: Sedment ...	+·9±·35	+2·57	+3·2±1·55	+ 2·06	+2·6±·86	+3·02	+·6±·55	+1·09
	XII	Congo Negro	+·3±·30	+1·00	+·8±1·36	+·59	—	—	+·1±·62	+·16
C. American Series	II	Peruvian	+·5±·16	+3·09	+5·7±·69	+ 8·26	+1·3±·94	+1·38	−·9±·26	−3·46
	I	Eskimo	−·3±·28	−1·07	+6·8±1·25	+ 5·44	—	—	−1·6±·50	−3·20
	III	North American Indian ...	−·5±·32	−1·56	+1·9±1·49	+ 1·28	—	—	−·8±·51	−1·57
D. Oceanic Series	XVI	Australian: "Other" regions	+·4±·23	+1·74	+9·9±·91	+10·88	+3·5±·68	+5·15	−·2±·36	−·56
	VI	Kanaka	+·4±·25	+1·60	+8·1±1·39	+ 5·83	+3·1±·68	+4·56	−1·4±·37	−3·79
	II	Moriori	−·6±·32	−1·88	+8·8±1·76	+ 5·00	—	—	−·2±·57	−·35
	XI	Papuan	−·4±·23	−1·74	+5·6±1·12	+ 5·00	+1·5±·68	+2·21	−1·0±·52	−1·92
	I	Maori	+·9±·31	+2·90	+6·4±1·54	+ 4·16	—	—	−·7±·58	−1·21
	III	New British	−·8±·34	−2·35	+7·3±1·71	+ 4·23	—	—	−·1±·56	−·18
E. Asiatic Series	XII	Javanese	+·4±·23	+1·74	+5·2±1·02	+ 5·10	+2·0±·64	+3·13	−·7±·48	−1·46
	XXIII	Hindu: N.E. India ...	+·6±·23	+2·61	+4·2±1·17	+ 3·60	+4·2±·87	+4·83	+·1±·35	+·29
	II	Burmese	+·7±·27	+2·59	+3·3±1·15	+ 2·87	—	—	+·7±·52	+1·35
	XV	Sarawak (Natives) ...	−·1±·33	−·30	+2·0±1·57	+ 1·27	—	—	—	—
	IX	Andamanese	+·1±·34	+·29	+1·4±1·90	+·74	—	—	−·8±·63	−1·27
		Average difference for all Series ...	+0·11		+5·21		+2·65		−·43	

* The series for each continental area are arranged in order according to the size of the sex ratio for the simotic index.

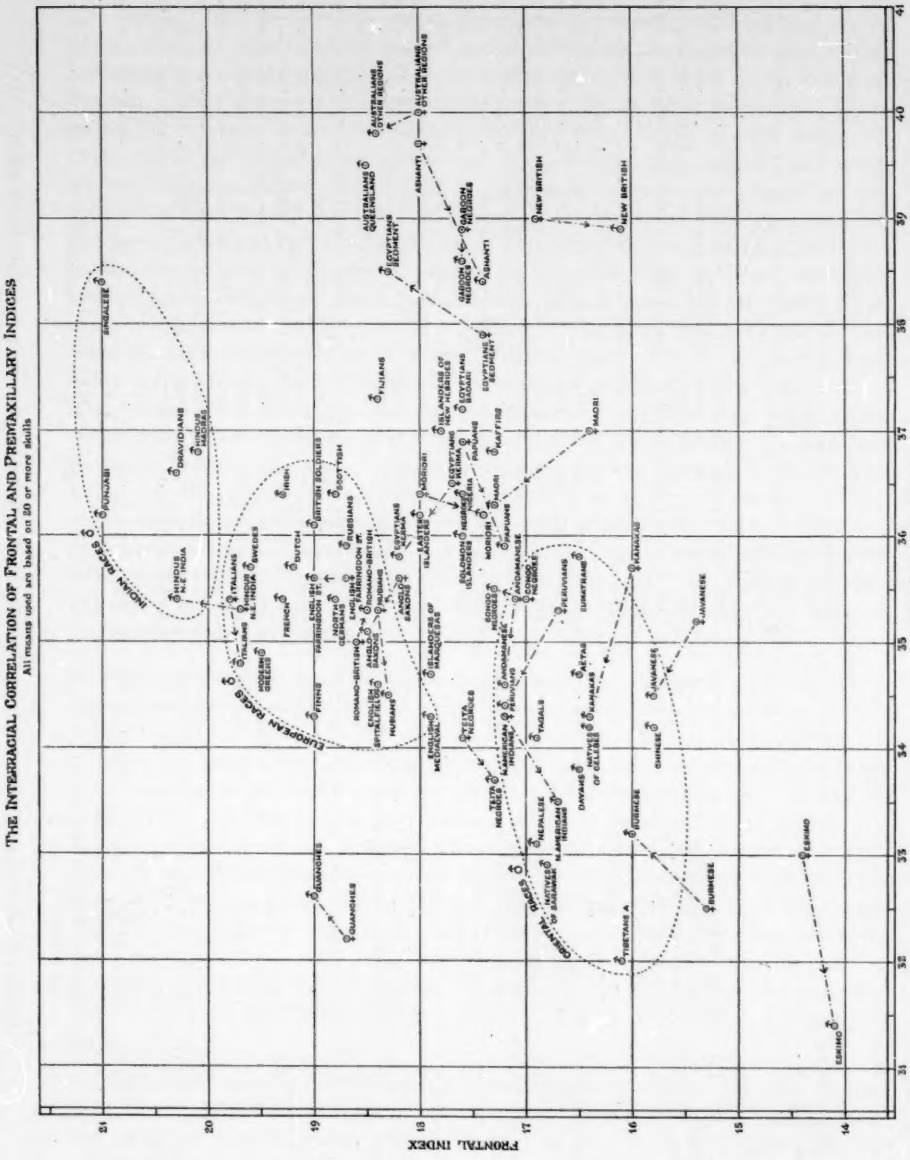

THE INTERRACIAL CORRELATION OF FRONTAL AND PREMAXILLARY INDICES
All means used are based on 30 or more skulls
INDIAN RACES
EUROPEAN RACES
ESKIMO
PREMAXILLARY INDEX
FRONTAL INDEX

Fig. 2.

the constants in Tables II (p. 226) and III (p. 229) that the races of Asia and Europe show a slight tendency to be less differentiated sexually than those of the other regions, while the Oceanic races tend to show the greatest sexual differences.

(5) *Sexual and Racial Differences in Variability.* Table IV gives the standard deviations and coefficients of variation for all the racial distributions of absolute measurements made up by 50 or more individuals: Table V gives the standard deviations of the indices for all the distributions made up by 30 or more individuals. The sexual differences in variability may be considered first. Few comparisons can be made in the case of the absolute measurements, but the data suggest that female variation for the characters considered shows a slight tendency to be greater than male variation. Combining all the measurements, there are 18 cases in which the male coefficient of variation exceeds the female and 26 in which the position is reversed. The most significant difference is only 3·4 times its probable error, and among the four cases showing this ratio greater than 3·0 the female index is the greater in three. It is curious that the position appears to be different when the indices are considered. In 44 comparisons the male standard deviation is the greater and for the remaining 14 the female constant is greater than the male. In six cases the sexual difference in variability exceeds 3·0 times its probable error, the greatest ratio being 6·4, and for all these the male constant is the greater. The dominance of male variation for the indices may also be appreciated by comparing the weighted mean standard deviations given in Table V. Out of all possible sexual comparisons between these which can be made for the four indices and five continental areas, there are 16 cases in which the male constant is in excess of the female and only four for which the position is reversed. We must conclude that for these measurements of shape male variation shows a distinct tendency to be greater than female, which for the measurements of size, judging by coefficients of variation, there is either sexual equality in variability, or females show a slight tendency to be more variable than males of the same race. The material is obviously too meagre to investigate whether there are any differences in the preponderance of variability of one sex over the other between different continental areas. Racial comparisons of the constants of variability may be considered next and it will be sufficient if we confine our attention to the male indices in this connection. The series in Table V are arranged in order of their standard deviations for the frontal index in the case of each continental area. The range of values for each area is quite wide and there is no difficulty in selecting pairs belonging to the same group which differ quite significantly. This may not indicate in every case that the racial populations from which the samples were drawn really differ in variability for this particular character, as several of the samples may not have been randomly chosen owing to reasons of which we are ignorant. If the distributions of the standard deviations for the five continental areas are compared it will be seen that they are all overlapping and the inter-group means—excluding the American which can only be based on three series—are almost identical. The same is found in the case of the premaxillary index, and so few series are available for the rhinal index that

Fig. 2.

TABLE IV.

Variabilities of Absolute Measurements for Distributions of Fifty or more Skulls.

Continental Areas	Serial No.	Race	Sex	Standard Deviations							
				Frontal		Simotic		Rhinal		Premaxillary	
				Chord	Subtense	Chord	Subtense	Chord	Subtense	Chord	Subtense
A. Europe	II	Romano-British	♂	4·04 ± ·21 (86)	2·33 ± ·12	1·79 ± ·09 (82)	1·27 ± ·07	—	—	4·62 ± ·28 (64)	2·92 ± ·17
	III	Anglo-Saxon	„	3·78 ± ·19 (92)	2·23 ± ·11	1·94 ± ·11 (77)	1·17 ± ·06	—	—	5·22 ± ·29 (76)	3·42 ± ·19
	V	English: Farringdon St. ...	„	3·88 ± ·21 (75)	1·94 ± ·11	1·92 ± ·10 (81)	1·10 ± ·06	—	—	6·12 ± ·40 (53)	3·27 ± ·27
	VI	English: Spitalfields ...	„	3·55 ± ·11 (248)	1·83 ± ·06	1·90 ± ·08 (126)	1·21 ± ·05	4·54 ± ·30 (50)	2·37 ± ·16	3·93 ± ·22 (73)	2·53 ± ·14
	IX	British (Soldiers)	„	4·81 ± ·22 (113)	2·41 ± ·11	1·59 ± ·07 (111)	1·15 ± ·05	4·74 ± ·25 (84)	3·18 ± ·17	5·03 ± ·24 (101)	2·58 ± ·12
	VII	Modern Irish	„	3·28 ± ·19 (65)	2·57 ± ·15	1·68 ± ·10 (65)	1·07 ± ·06	—	—	3·64 ± ·23 (57)	2·63 ± ·17
	XII	Italian	„	4·25 ± ·20 (100)	2·38 ± ·11	2·12 ± ·10 (100)	1·15 ± ·05	4·60 ± ·30 (52)	2·64 ± ·17	4·80 ± ·24 (94)	3·23 ± ·16
	XVI	Dutch	„	4·28 ± ·29 (51)	2·79 ± ·19	1·99 ± ·13 (51)	1·19 ± ·08	—	—	—	—
	V	English: Farringdon St. ...	♀	3·87 ± ·21 (75)	2·23 ± ·12	1·74 ± ·09 (76)	·99 ± ·05	—	—	4·33 ± ·29 (51)	2·38 ± ·16
	VI	English: Spitalfields ...	„	3·31 ± ·16 (94)	1·68 ± ·08	—	—	—	—	—	—
				Coefficients of Variation							
A. Europe	II	Romano-British	♂	4·10 ± ·21	12·73 ± ·67	19·26 ± 1·05	27·50 ± 1·55	—	—	4·88 ± ·29	8·91 ± ·54
	III	Anglo-Saxon	„	3·86 ± ·19	12·26 ± ·62	21·33 ± 1·21	24·33 ± 1·40	—	—	5·49 ± ·30	10·24 ± ·57
	V	English: Farringdon St. ...	„	3·95 ± ·22	10·41 ± ·58	20·87 ± 1·15	23·91 ± 1·33	—	—	6·62 ± ·44	8·79 ± ·58
	VI	English: Spitalfields ...	„	3·62 ± ·11	10·19 ± ·31	19·59 ± ·86	26·89 ± 1·21	8·37 ± ·57	10·77 ± ·73	4·21 ± ·24	7·86 ± ·44
	IX	British (Soldiers)	„	4·92 ± ·22	13·01 ± ·60	17·85 ± ·83	24·37 ± 1·17	8·72 ± ·46	13·20 ± ·70	5·40 ± ·26	7·67 ± ·37
	VII	Modern Irish	„	3·31 ± ·20	13·37 ± ·81	19·58 ± 1·20	23·20 ± 1·44	—	—	3·89 ± ·25	7·69 ± ·49
	XII	Italian	„	4·33 ± ·21	12·34 ± ·60	21·81 ± 1·09	25·48 ± 1·29	8·57 ± ·57	10·91 ± ·73	5·08 ± ·25	9·89 ± ·49
	XVI	Dutch	„	4·31 ± ·29	14·63 ± 1·00	22·15 ± 1·55	25·34 ± 1·80	—	—	—	—
	V	English: Farringdon St. ...	♀	4·11 ± ·23	12·68 ± ·71	19·12 ± 1·07	24·75 ± 1·43	—	—	4·99 ± ·33	7·74 ± ·52
	VI	English: Spitalfields ...	„	3·54 ± ·17	9·78 ± ·49	—	—	—	—	—	—

				Standard Deviations							
B. Africa	II	Nubian	♂	3·25 ± ·19 (69)	2·43 ± ·14	2·18 ± ·14 (55)	1·05 ± ·07	—	—	—	—
	V	Egyptian: Kerma ...	,,	3·70 ± ·16 (117)	2·08 ± ·09	1·92 ± ·09 (99)	1·19 ± ·06	—	—	4·53 ± ·21 (103)	3·31 ± ·16
	VI	Nigerian (Natives) ...	,,	4·04 ± ·27 (52)	1·83 ± ·12	—	—	—	—	—	—
	VII	Gaboon Negro	,,	3·61 ± ·20 (76)	2·55 ± ·14	2·63 ± ·15 (71)	·98 ± ·06	—	—	4·76 ± ·29 (63)	2·86 ± ·17
	X	Negro: Teita Hills ...	,,	3·78 ± ·24 (55)	2·38 ± ·13	—	—	—	—	—	—
	XII	Congo Negro	,,	3·52 ± ·22 (60)	2·28 ± ·14	2·08 ± ·13 (60)	·86 ± ·05	—	—	—	—
	XV	Kaffir	,,	3·94 ± ·26 (54)	2·66 ± ·17	2·43 ± ·16 (54)	1·11 ± ·07	—	—	4·87 ± ·33 (51)	3·35 ± ·22
	II	Nubian	♀	3·84 ± ·23 (61)	2·12 ± ·13	—	—	—	—	—	—
	V	Egyptian: Kerma ...	,,	3·18 ± ·16 (94)	2·16 ± ·11	2·04 ± ·11 (74)	1·07 ± ·06	—	—	4·18 ± ·22 (83)	3·15 ± ·16
	VII	Gaboon Negro	,,	4·58 ± ·28 (60)	1·99 ± ·12	2·48 ± ·16 (55)	·83 ± ·05	—	—	5·77 ± ·38 (51)	3·75 ± ·25
	X	Negro: Teita Hills ...	,,	3·12 ± ·19 (62)	1·99 ± ·12	—	—	—	—	—	—

				Coefficients of Variation							
B. Africa	II	Nubian	♂	3·45 ± ·20	13·89 ± ·81	18·60 ± 1·24	25·10 ± 1·71	—	—	—	—
	V	Egyptian: Kerma ...	,,	3·85 ± ·17	11·88 ± ·53	18·48 ± ·92	31·24 ± 1·64	—	—	4·78 ± ·23	9·76 ± ·46
	VI	Nigerian (Natives) ...	,,	4·06 ± ·27	10·44 ± ·70	—	—	—	—	—	—
	VII	Gaboon Negro	,,	3·64 ± ·20	14·66 ± ·82	28·27 ± 1·72	35·59 ± 2·26	—	—	4·97 ± ·30	7·75 ± ·47
	X	Negro: Teita Hills ...	,,	3·82 ± ·25	13·84 ± ·91	—	—	—	—	—	—
	XII	Congo Negro	,,	3·57 ± ·22	13·32 ± ·83	21·88 ± 1·41	34·40 ± 2·36	—	—	—	—
	XV	Kaffir	,,	3·82 ± ·25	14·93 ± ·99	28·54 ± 2·00	29·53 ± 2·94	—	—	5·05 ± ·34	9·47 ± ·64
	II	Nubian	♀	4·12 ± ·25	12·47 ± ·77	—	—	—	—	—	—
	V	Egyptian: Kerma ...	,,	3·45 ± ·17	13·27 ± ·66	20·64 ± 1·19	33·50 ± 2·06	—	—	4·61 ± ·24	9·53 ± ·50
	VII	Gaboon Negro	,,	4·83 ± ·30	11·89 ± ·74	25·57 ± 1·75	33·30 ± 2·37	—	—	6·36 ± ·43	10·67 ± ·72
	X	Negro: Teita Hills ...	,,	3·30 ± ·20	11·97 ± ·74	—	—	—	—	—	—

TABLE IV—(*continued*).

Continental Areas	Serial No.	Race	Sex	Standard Deviations							
				Frontal		Simotic		Rhinal		Premaxillary	
				Chord	Subtense	Chord	Subtense	Chord	Subtense	Chord	Subtense
C. America	I	Eskimo	♂	3·45±·23 (51)	2·06±·14	1·84±·12 (50)	·95±·06	—	—	—	—
	II	Peruvian	,,	3·31±·12 (171)	1·98±·07	1·78±·07 (166)	·86±·03	—	—	4·33±·17 (143)	3·05±·12
	II	Peruvian	♀	3·44±·15 (114)	1·97±·09	1·44±·07 (107)	·79±·04	—	—	4·32±·21 (97)	2·99±·14
				Coefficients of Variation							
C. America	I	Eskimo	♂	3·48±·23	14·72±1·00	32·30±2·40	39·61±3·06	—	—	—	—
	II	Peruvian	,,	3·41±·12	11·87±·44	19·37±·73	22·65±·88	—	—	4·35±·17	8·92±·37
	II	Peruvian	♀	3·69±·17	12·72±·58	15·94±·75	24·56±1·20	—	—	4·58±·22	8·98±·44
				Standard Deviations							
D. Oceania	I	Maori: New Zealand ...	♂	4·66±·24 (89)	2·09±·11	1·51±·08 (87)	1·04±·05	—	—	5·46±·30 (76)	3·43±·19
	VI	Kanaka	,,	3·66±·21 (67)	2·33±·14	1·55±·09 (63)	·91±·05	—	—	5·04±·33 (54)	3·00±·19
	XI	Papuan	,,	4·02±·22 (76)	1·93±·11	2·05±·11 (78)	·86±·05	—	—	5·77±·34 (65)	2·82±·17
	XVI	Australian: "Other" regions	,,	3·95±·17 (121)	2·45±·11	1·78±·08 (114)	·95±·04	4·31±·41 (50)	2·51±·17	5·08±·24 (103)	2·98±·14
	VI	Kanaka	♀	4·04±·25 (60)	2·02±·12	1·61±·10 (60)	·90±·06	—	—	5·14±·32 (57)	3·04±·19
	XVI	Australian: "Other" regions	,,	3·47±·19 (80)	1·96±·10	1·93±·11 (74)	·91±·05	—	—	5·58±·32 (71)	3·22±·18
				Coefficients of Variation							
D. Oceania	I	Maori: New Zealand ...	♂	4·70±·24	12·28±·63	20·93±1·12	33·46±1·89	—	—	5·52±·30	9·68±·53
	VI	Kanaka	,,	3·74±·22	14·54±·87	20·96±1·31	33·57±2·23	—	—	5·13±·33	8·94±·58
	XI	Papuan	,,	4·13±·23	11·54±·64	25·30±1·45	26·80±1·55	—	—	5·96±·35	8·16±·49
	XVI	Australian: "Other" regions	,,	3·39±·15	13·08±·58	19·16±·89	23·08±1·08	7·48±·51	13·15±·90	5·47±·26	8·05±·38
	VI	Kanaka	♀	4·36±·27	13·46±·84	22·39±1·45	42·81±3·08	—	—	5·59±·35	9·31±·59
	XVI	Australian: "Other" regions	,,	3·62±·19	11·33±·61	21·89±1·27	30·28±1·83	—	—	6·29±·36	9·08±·52

	XVI	Australian: "Other" regions	,,	3·62 ± ·19	11·33 ± ·61	21·89 ± 1·27	30·28 ± 1·83	—	—	5·59 ± ·35 6·29 ± ·36	9·31 ± ·59 9·05 ± ·52

				Standard Deviations							
E. Asia	I	Chinese	♂	4·07 ± ·17 (136)	2·33 ± ·10	1·94 ± ·08 (137)	·94 ± ·04	4·99 ± ·23 (109)	2·71 ± ·12	4·51 ± ·20 (121)	3·27 ± ·14
	II	Burmese	,,	4·79 ± ·25 (84)	2·51 ± ·13	2·32 ± ·12 (82)	1·01 ± ·05	5·06 ± ·32 (57)	2·57 ± ·16	4·97 ± ·28 (70)	2·98 ± ·17
	XII	Javanese	,,	4·52 ± ·16 (176)	2·20 ± ·08	1·98 ± ·07 (172)	1·00 ± ·04	4·60 ± ·18 (148)	2·57 ± ·10	4·70 ± ·17 (166)	3·53 ± ·13
	XIII	Sumatran	,,	3·43 ± ·22 (56)	1·56 ± ·10	2·27 ± ·14 (56)	1·01 ± ·06	5·30 ± ·35 (52)	2·87 ± ·19	4·71 ± ·32 (50)	3·45 ± ·23
	XIV	Celebes (Natives)	,,	4·11 ± ·27 (52)	2·36 ± ·16	2·32 ± ·15 (52)	·85 ± ·06	—	—	—	—
	XV	Sarawak (Natives) ...	,,	4·01 ± ·26 (55)	2·16 ± ·14	1·86 ± ·12 (55)	·87 ± ·06	—	—	—	—
	XVI	Dayak	,,	4·21 ± ·27 (55)	2·12 ± ·14	2·19 ± ·15 (50)	·88 ± ·06	—	—	—	—
	XXI	Punjabi	,,	3·70 ± ·18 (93)	2·65 ± ·13	1·85 ± ·09 (92)	1·27 ± ·06	—	—	4·73 ± ·25 (79)	3·27 ± ·18
	XXIII	Hindu: N.E. India ...	,,	3·92 ± ·14 (179)	2·38 ± ·08	1·68 ± ·06 (173)	1·10 ± ·04	4·26 ± ·23 (78)	2·42 ± ·13	5·01 ± ·19 (163)	3·52 ± ·13
	XXIII	Hindu: N.E. India ...	♀	3·67 ± ·25 (50)	2·28 ± ·15	—	—	—	—	—	—

				Coefficients of Variation							
E. Asia	I	Chinese	♂	4·23 ± ·17	15·43 ± ·65	25·57 ± 1·11	39·08 ± 1·82	9·15 ± ·42	15·92 ± ·75	4·54 ± ·20	9·63 ± ·42
	II	Burmese	,,	4·87 ± ·25	15·96 ± ·85	26·04 ± 1·46	33·73 ± 1·97	8·79 ± ·56	15·22 ± ·98	4·89 ± ·28	8·85 ± ·51
	XII	Javanese	,,	4·66 ± ·17	14·31 ± ·52	22·75 ± ·87	35·65 ± 1·45	8·11 ± ·32	14·83 ± ·59	4·67 ± ·17	10·21 ± ·38
	XIII	Sumatran	,,	3·55 ± ·23	9·82 ± ·63	27·07 ± 1·85	37·25 ± 2·68	9·54 ± ·64	15·97 ± 1·08	4·73 ± ·32	9·71 ± ·66
	XIV	Celebes (Natives)	,,	4·15 ± ·27	14·46 ± ·98	26·70 ± 1·89	31·46 ± 2·28	—	—	—	—
	XV	Sarawak (Natives) ...	,,	4·16 ± ·27	13·32 ± ·87	23·29 ± 1·58	33·29 ± 2·37	—	—	—	—
	XVI	Dayak	,,	4·36 ± ·28	13·34 ± ·87	25·79 ± 1·85	32·56 ± 2·42	—	—	—	—
	XXI	Punjabi	,,	3·86 ± ·19	13·14 ± ·66	20·36 ± 1·05	28·13 ± 1·51	—	—	5·04 ± ·27	9·70 ± ·53
	XXIII	Hindu: N.E. India ...	,,	4·11 ± ·15	12·26 ± ·44	19·32 ± ·73	26·90 ± 1·04	7·47 ± ·41	10·65 ± ·58	5·30 ± ·20	10·54 ± ·40
	XXIII	Hindu: N.E. India ...	♀	4·00 ± ·27	12·61 ± ·86	—	—	—	—	—	—

TABLE V. *Standard Deviations of the Indices of Facial Flattening**.

Continental Area	Serial No.	Race	Sex	Frontal Index	Simotic Index	Rhinal Index	Premaxillary Index
A. Europe	VI	English: Spitalfields ...	♂	1·70 ± ·05 (248)	11·79 ± ·50 (126)	5·15 ± ·35 (50)	3·08 ± ·17 (73)
	IV	English: Mediaeval ...	,,	1·87 ± ·14 (43)	12·79 ± 1·02 (36)	—	3·67 ± ·30 (35)
	XX	Russian	,,	2·03 ± ·16 (35)	11·72 ± ·93 (36)	—	2·66 ± ·22 (33)
	III	Anglo-Saxon	,,	2·08 ± ·10 (92)	9·29 ± ·51 (77)	—	3·65 ± ·20 (76)
	XI	French	,,	2·11 ± ·15 (46)	8·98 ± ·65 (43)	—	4·11 ± ·31 (39)
	II	Romano-British	,,	2·18 ± ·11 (86)	12·51 ± ·66 (82)	5·56 ± ·48 (30)	2·96 ± ·18 (64)
	V	English: Farringdon St....	,,	2·19 ± ·12 (75)	12·75 ± ·68 (81)	6·16 ± ·48 (37)	3·67 ± ·24 (53)
	XVIII	Modern Greek	,,	2·26 ± ·19 (33)	10·02 ± ·85 (32)	—	—
	VIII	Modern Scottish	,,	2·30 ± ·19 (35)	9·40 ± ·79 (32)	—	3·14 ± ·27 (31)
	X	North German	,,	2·31 ± ·20 (32)	11·90 ± 1·04 (30)	—	3·65 ± ·31 (31)
	XII	Italian	,,	2·33 ± ·11 (100)	11·40 ± ·54 (100)	5·78 ± ·38 (52)	3·44 ± ·17 (94)
	XIV	Swedish	,,	2·36 ± ·21 (30)	—	—	—
	VII	Modern Irish	,,	2·37 ± ·14 (65)	11·70 ± ·69 (65)	5·35 ± ·39 (44)	2·65 ± ·17 (57)
	IX	British (Soldiers)	,,	2·45 ± ·11 (113)	13·13 ± ·59 (111)	6·05 ± ·31 (84)	3·12 ± ·15 (101)
	XVI	Dutch	,,	2·49 ± ·17 (51)	11·15 ± ·74 (51)	—	3·32 ± ·24 (45)
		Weighted Mean ($\bar{\sigma}$) ...	,,	2·14 [15]	11·64 [14]	5·77 [6]	3·32 [13]
	III	Anglo-Saxon	♀	1·59 ± ·11 (45)	9·32 ± ·69 (41)	—	2·82 ± ·21 (40)
	VI	English: Spitalfields ...	,,	1·76 ± ·09 (94)	11·63 ± ·97 (31)	—	—
	II	Romano-British	,,	2·14 ± ·15 (45)	10·73 ± ·79 (42)	—	3·56 ± ·29 (39)
	V	English: Farringdon St....	,,	2·15 ± ·12 (75)	10·19 ± ·56 (76)	5·27 ± ·37 (47)	3·99 ± ·27 (51)
	XII	Italian	,,	2·64 ± ·22 (32)	10·67 ± ·91 (31)	—	—
		Weighted Mean ($\bar{\sigma}$) ...	,,	2·00 [5]	10·42 [5]	5·27 [1]	3·53 [3]
B. Africa	IX	Ashanti	♂	1·79 ± ·15 (33)	10·66 ± ·87 (34)	—	3·15 ± ·27 (30)
	VI	Negro: Nigeria	,,	1·82 ± ·12 (52)	9·66 ± ·66 (49)	—	3·60 ± ·25 (46)
	III	Egyptian: Badari ...	,,	2·09 ± ·18 (32)	7·07 ± ·58 (34)	—	3·37 ± ·28 (34)
	VII	Gaboon Negro	,,	2·18 ± ·12 (76)	8·70 ± ·49 (71)	4·18 ± ·28 (49)	3·27 ± ·20 (63)
	XII	Congo Negro	,,	2·21 ± ·14 (60)	8·23 ± ·51 (60)	—	4·26 ± ·31 (42)
	X	Teita Negro	,,	2·24 ± ·14 (55)	7·69 ± ·59 (39)	—	2·71 ± ·23 (31)
	IV	Egyptian: Sedment ...	,,	2·33 ± ·18 (40)	10·01 ± ·79 (37)	—	3·38 ± ·27 (37)
	V	Egyptian: Kerma ...	,,	2·41 ± ·11 (117)	9·19 ± ·44 (99)	—	3·62 ± ·17 (103)
	XV	Kaffir	,,	2·43 ± ·16 (54)	11·94 ± ·78 (54)	—	3·63 ± ·24 (51)
	II	Nubian	,,	2·49 ± ·14 (69)	8·09 ± ·52 (55)	—	3·62 ± ·27 (40)
		Weighted Mean ($\bar{\sigma}$) ...	,,	2·25 [10]	9·31 [10]	4·18 [1]	3·52 [10]
	XII	Congo Negro	♀	1·87 ± ·16 (30)	—	—	—
	II	Nubian	,,	1·90 ± ·12 (61)	6·15 ± ·48 (38)	—	2·56 ± ·21 (35)
	IV	Egyptian: Sedment ...	,,	2·00 ± ·17 (30)	—	—	—
	VII	Gaboon Negro	,,	2·01 ± ·12 (60)	8·42 ± ·54 (55)	3·96 ± ·29 (44)	3·24 ± ·22 (51)
	IX	Ashanti	,,	2·08 ± ·18 (31)	10·96 ± ·95 (30)	—	—
	X	Teita Negro	,,	2·09 ± ·13 (62)	7·16 ± ·49 (49)	4·26 ± ·37 (31)	3·69 ± ·29 (37)
	V	Egyptian: Kerma ...	,,	2·19 ± ·11 (94)	9·49 ± ·53 (74)	—	3·00 ± ·16 (83)
		Weighted Mean ($\bar{\sigma}$) ...	,,	2·05 [7]	8·58 [5]	4·09 [2]	3·12 [4]

* The numbers in round brackets denote the numbers of skulls on which the standard deviations are based and none was calculated for series made up by fewer than 30 specimens. The weighted mean $\bar{\sigma}$'s for the different continental areas were found from the formula $\bar{\sigma} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + \dots}{n_1 + n_2 + \dots}}$, and the numbers in square brackets denote the number of series on which they are based. The series for a particular continental area are arranged in order of the standard deviations of the frontal indices.

TABLE V—(*continued*).

Continental Area	Serial No.	Race	Sex	Frontal Index	Simotic Index	Rhinal Index	Premaxillary Index
C. America	III	North American Indian ...	♂	1·74±·12 (45)	11·31±·80 (46)	—	2·93±·22 (41)
	II	Peruvian	„	1·81±·07 (171)	9·51±·35 (166)	4·84±·35 (43)	3·00±·12 (143)
	I	Eskimo	„	1·96±·13 (51)	8·86±·60 (50)	—	3·23±·23 (45)
		Weighted Mean ($\bar{\sigma}$) ...	„	1·83 [3]	9·73 [3]	4·84 [1]	3·03 [3]
	II	Peruvian	♀	2·09±·09 (114)	7·36±·34 (107)	5·25±·56 (20)	2·96±·14 (97)
	I	Eskimo	„	1·72±·15 (30)	—	—	—
		Weighted Mean ($\bar{\sigma}$) ...	„	2·02 [2]	7·36 [1]	5·25 [1]	2·96 [1]
D. Oceania	II	Moriori	♂	1·84±·13 (43)	11·89±·88 (42)	—	2·88±·22 (39)
	XI	Papuan	„	1·84±·10 (76)	10·04±·54 (78)	5·06±·35 (47)	3·57±·21 (65)
	III	New British	„	1·85±·14 (39)	10·54±·81 (39)	3·46±·38 (33)	2·40±·19 (35)
	V	Easter Island	„	1·93±·10 (41)	12·24±·88 (44)	—	3·37±·27 (36)
	X	Fijian	„	1·98±·17 (30)	10·85±·95 (30)	—	—
	I	Maori	„	2·01±·12 (89)	12·20±·62 (87)	5·11±·40 (38)	3·46±·19 (76)
	VI	Kanaka: Sandwich Islands	„	2·09±·12 (67)	11·81±·71 (63)	5·39±·38 (47)	2·80±·18 (54)
	IX	New Hebridean	„	2·23±·18 (36)	10·13±·83 (34)	—	3·74±·31 (33)
	XV	Australian: Queensland ...	„	2·53±·20 (36)	8·98±·71 (36)	—	3·41±·28 (35)
	XVI	Australian: "Other" regions	„	2·99±·13 (121)	10·04±·45 (114)	4·62±·31 (50)	3·60±·17 (103)
		Weighted Mean ($\bar{\sigma}$) ...	„	2·26 [10]	11·22 [10]	4·83 [5]	3·34 [9]
	XI	Papuan	♀	1·86±·13 (49)	8·45±·58 (49)	3·95±·32 (34)	4·03±·30 (40)
	VI	Kanaka: Sandwich Islands	„	2·05±·13 (60)	11·05±·68 (60)	3·81±·30 (37)	2·98±·19 (57)
	XVI	Australian: "Other" regions	„	1·94±·10 (80)	8·26±·46 (74)	4·22±·39 (27)	3·41±·19 (71)
		Weighted Mean ($\bar{\sigma}$) ...	„	1·96 [3]	9·31 [3]	3·97 [3]	3·43 [3]
E. Asia	XIII	Sumatran	♂	1·58±·10 (56)	10·14±·65 (56)	5·87±·39 (52)	3·61±·24 (50)
	XVIII	Aëta	„	1·64±·14 (33)	—	—	—
	XVII	Tagal	„	1·65±·14 (31)	—	—	—
	IX	Andamanese	„	1·83±·16 (31)	—	—	—
	XVI	Dayak	„	2·00±·13 (55)	9·85±·67 (50)	4·46±·32 (44)	3·32±·23 (46)
	XV	Sarawak (Natives) ...	„	2·04±·13 (55)	10·05±·65 (55)	—	4·56±·33 (43)
	V	Tibetan *A*	„	2·06±·16 (37)	8·33±·67 (35)	5·71±·47 (34)	3·21±·26 (36)
	XXII	Hindu: Madras	„	2·06±·14 (49)	10·44±·73 (46)	—	3·44±·24 (48)
	XII	Javanese	„	2·13±·08 (176)	9·75±·36 (172)	4·82±·19 (148)	3·83±·14 (166)
	XXIV	Dravidian	„	2·15±·17 (36)	10·46±·72 (35)	—	3·33±·29 (31)
	I	Chinese	„	2·19±·09 (136)	10·79±·44 (137)	5·18±·24 (109)	3·51±·15 (121)
	XIV	Celebes (Natives)	„	2·19±·15 (52)	7·44±·49 (52)	4·29±·31 (45)	3·42±·23 (49)
	II	Burmese	„	2·24±·12 (84)	8·95±·47 (82)	3·81±·24 (57)	2·99±·17 (70)
	XIX	Singalese	„	2·24±·18 (34)	10·37±·85 (34)	—	—
	XXIII	Hindu: N.E. India ...	„	2·35±·08 (179)	11·17±·41 (173)	5·02±·27 (78)	3·44±·13 (163)
	VII	Nepalese	„	2·56±·18 (46)	14·48±1·01 (47)	5·79±·44 (39)	2·97±·21 (44)
	XXI	Punjabi	„	2·87±·14 (93)	12·94±·64 (92)	7·64±·56 (43)	3·67±·20 (79)
		Weighted Mean ($\bar{\sigma}$) ...	„	2·20 [17]	10·60 [14]	5·20 [10]	3·54 [13]
	XII	Javanese	♀	1·75±·14 (34)	7·68±·64 (34)	4·72±·41 (30)	3·68±·32 (30)
	II	Burmese	„	1·99±·15 (38)	8·49±·67 (37)	—	2·75±·24 (30)
	XXIII	Hindu: N.E. India ...	„	2·07±·14 (50)	10·46±·72 (48)	—	2·96±·21 (46)
		Weighted Mean ($\bar{\sigma}$) ...	„	1·96 [3]	9·13 [3]	4·72 [1]	3·13 [3]

no conclusion can be drawn. The position is almost the same for the simotic index except that the group mean for the African races is appreciably lower than the others. It may be seen from Fig. 1 that the lowest simotic indices found are for African races. We can only conclude that the variabilities of the indices of facial flattening for different races are not peculiar to continental areas and that there is no evidence that the variations of advanced and primitive peoples are differentiated in this respect.

(6) *Intra- and Interracial Correlations of the Measurements of Facial Flattening.* All four of the indices considered in the present paper are of the same type, as each is formed by expressing a "median" subtense as a percentage of a corresponding transverse breadth. It might be anticipated that for both intra- and interracial samples they would be quite highly correlated with one another, and this could be supposed an expression of the fact that they might all be presumed to measure the same feature of the cranium (viz. the transverse flattening of the facial skeleton) in rather different ways. For an intraracial sample positive correlations are to be expected between all pairs of the absolute measurements, owing to the effect of a common size factor*, and it would not be surprising to find that some pairs of the breadths and some pairs of the subtenses show high positive correlations. The correlations actually found do not coincide with all our anticipations in the present case. Table VI gives 72 male and female intraracial correlation coefficients for the longest series available. They range from $-\cdot176$ to $+\cdot434$ and there are only 29 values (all positive) which can be considered to differ significantly from zero. The frontal index is seen to be quite uncorrelated with the simotic, no significant correlation being found for any series, while the closest association is that between the simotic and rhinal indices which both measure the prominence of the nasal bones, all the values being significant in this case. Both frontal and premaxillary indices show a more marked, but still low, degree of association with the rhinal than with the simotic index. We must conclude that the last, which gives a measure of the prominence of the root of the nasal bones, is practically uncorrelated in the individual with other measurements of facial flattening which do not involve the nasal bones. The same might be true for an index confined to the lower extremities of these bones, the rhinal not being of that type since the breadth from which it is calculated is between points on the malar and maxillary bones. But the correlations between the frontal and premaxillary indices are also of a low order and the supposition that all our indices were measuring the same feature of the facial skeleton in rather different ways is seen to be entirely belied by the facts. The intraracial correlations of the absolute measurements will clearly be of interest and these are given in Table VII for the male Javanese series—the longest available—for which the indicial correlations in Table VI are quite typical. As might have been anticipated, all the correlations

* See the standard paper on cranial correlations by Professor Karl Pearson and Miss Adelaide G. Davin: "On the Biometric Constants of the Human Skull," *Biometrika*, Vol. XVI. (1924), pp. 328—363.

TABLE VI.

Intraracial Correlations of the Indices of Facial Flattening.*

Race	Sex	$r_{F,S}$ (Frontal and Simotic)	$r_{F,R}$ (Frontal and Rhinal)	$r_{F,P}$ (Frontal and Premaxillary)	$r_{S,R}$ (Simotic and Rhinal)	$r_{S,P}$ (Simotic and Premaxillary)	$r_{R,P}$ (Rhinal and Premaxillary)
British (Soldiers)	♂	+·083±·064 (111)	*+·299±·067* (84)	*+·242±·063* (101)	*+·248±·069* (84)	*+·199±·065* (98)	*+·243±·073* (76)
Italian	„	−·089±·067 (101)	*+·279±·086* (52)	*+·203±·066* (96)	*+·393±·080* (51)	*+·237±·066* (95)	*+·288±·087* (50)
Chinese	„	+·083±·058 (134)	*+·301±·059* (108)	*+·178±·059* (121)	*+·381±·055* (109)	*+·242±·058* (121)	*+·331±·060* (100)
Hindu: N.E. India ...	„	+·076±·051 (173)	*+·434±·062* (78)	*+·243±·050* (163)	*+·373±·065* (78)	*+·169±·052* (158)	*+·371±·068* (73)
Javanese	„	+·064±·052 (165)	+·153±·054 (149)	*+·169±·051* (166)	*+·399±·047* (143)	+·119±·052 (163)	+·127±·056 (142)
Egyptian: Kerma ...	„	+·002±·069 (96)	—	+·074±·067 (99)	—	+·086±·071 (88)	—
Gaboon Negro	„	+·115±·078 (72)	−·069±·096 (49)	+·113±·083 (64)	*+·272±·089* (49)	+·093±·086 (60)	+·147±·102 (42)
Peruvian	„	+·137±·051 (166)	+·141±·101 (43)	*+·253±·053* (143)	*+·413±·085* (43)	+·020±·056 (143)	+·189±·106 (38)
Maori	„	+·173±·070 (88)	+·035±·111 (37)	−·052±·077 (76)	*+·351±·096* (38)	+·182±·075 (75)	−·038±·114 (35)
Australian: "Other" regions	„	−·038±·064 (111)	+·047±·096 (49)	+·082±·066 (102)	*+·282±·088* (50)	−·024±·067 (101)	*+·400±·086* (43)
Egyptian: Kerma ...	♀	−·176±·077 (72)	—	+·141±·074 (80)	—	−·061±·081 (69)	—
Gaboon Negro	„	−·144±·091 (53)	+·089±·101 (44)	+·007±·095 (50)	*+·426±·082* (44)	+·082±·099 (46)	+·101±·107 (39)
Peruvian	„	+·096±·065 (107)	—	+·156±·063 (96)	—	−·109±·070 (92)	—
Australian: "Other" regions	„	−·035±·079 (73)	—	+·093±·080 (70)	—	+·142±·080 (69)	—

* The coefficients in italics are the only ones which differ from zero by more than three times their probable errors.

in Table VII are positive. It has been shown by Pearson and Davin that nearly all the intraracial values found between pairs of absolute measurements are positive owing to the influence of a common size factor. This could be supposed to account for a number of correlations found for the Egyptian series with which they were dealing of the order +·1 to +·4†, but when the measurements compared also partly "covered" one another higher values were generally found‡. Among the 28 coefficients in Table VII there are 10 less than ·2, and most of these do not

† E.g. the coefficients found for the male Egyptian *E* series are +·3971±·0191 for *L* and *B*, +·3344±·0203 for *B* and *H*, +·2758±·0229 for *G'H* and *J*, and +·2358±·0218 for *NB* and *L*.

‡ E.g. the coefficients found for the male Egyptian *E* series are +·6341±·0137 for *B* and *U*, +·7059±·0117 for *NH* and *G'H*, and +·4170±·0199 for *B'* and *J*.

TABLE VII.

Intraracial Correlations of Absolute Measurements for the Male Javanese Series, all values based on the same 137 *Skulls.*

Characters	Sub. *IOW*	*SS*	Sub. *MOW*	Sub. *GB*	*IOW*	*SC*	*MOW*	*GB*
Sub. *IOW*	—	+·303±·052	+·396±·049	+·225±·055	+·445±·046	+·437±·047	+·366±·050	+·141±·056
SS	+·303±·052	—	+·367±·050	+·087±·057	+·192±·056	+·529±·042	+·071±·057	+·044±·058
Sub. *MOW*	+·396±·049	+·367±·050	—	+·259±·054	+·201±·055	+·091±·057	+·237±·054	+·151±·056
Sub. *GB*	+·225±·055	+·087±·057	+·259±·054	—	+·267±·054	+·047±·058	+·213±·055	+·143±·056
IOW	+·445±·046	+·192±·056	+·201±·055	+·267±·054	—	+·305±·052	+·450±·046	+·611±·036
SC	+·437±·047	+·529±·042	+·091±·057	+·047±·058	+·305±·052	—	+·222±·055	+·022±·058
MOW	+·366±·050	+·071±·057	+·237±·054	+·213±·055	+·450±·046	+·222±·055	—	+·418±·048
GB	+·141±·056	+·044±·058	+·151±·056	+·143±·057	+·611±·036	+·022±·058	+·418±·048	—

TABLE VIII.

Interracial Indicial Constants for Series made up by Fifteen or more Crania.*

Indices	Sex	Means	Standard Deviations	Correlations of Indices			
				Frontal	Simotic	Rhinal	Premaxillary
Frontal ...	♂	17·95± ·109 (76)	1·412±·077 (76)	—	+·699±·040 (75)	+·679±·054 (45)	+·320±·072 (70)
	♀	17·35± ·132 (37)	1·189±·093 (37)	—	+·574±·075 (36)	—	+·244±·110 (33)
Simotic ...	♂	40·40± ·649 (75)	8·338±·459 (75)	+·699±·040 (75)	—	+·868±·025 (45)	+·136±·079 (70)
	♀	32·58± ·832 (36)	7·399±·588 (36)	+·574±·075 (36)	—	—	−·032±·117 (33)
Rhinal ...	♂	35·97± ·414 (45)	4·113±·292 (45)	+·679±·054 (45)	+·868±·025 (45)	—	+·070±·100 (45)
	♀	32·64±1·024 (14)	5·679±·724 (14)	—	—	—	—
Premaxillary	♂	35·79± ·161 (70)	1·994±·114 (70)	+·320±·072 (70)	+·136±·079 (70)	+·070±·100 (45)	—
	♀	35·76± ·225 (33)	1·917±·159 (33)	+·244±·110 (33)	−·032±·117 (33)	—	—

* In calculating the interracial constants in Tables VIII—X no account was taken of the fact that the means are of unequal value since the numbers of individuals on which they are based differ widely. Since no system of weighting the means was employed, the usual formulae were applied to give the probable errors, and these would not be applicable, of course, if weighted means were used to determine the interracial constants. The procedure adopted is probably accurate enough for the purpose in view.

differ significantly from zero, while there are only two greater than ·5*. The different subtenses show no high correlations with one another and the simotic chord is lowly correlated with the other facial breadths. We are dealing, in fact, with characters which are but little associated intraracially, and the conception of there being a single factor of facial features which can be measured in the individual in slightly different ways by the different subtenses or indices considered is seen to be of no practical utility.

Since the indices we are considering were devised for the purpose of discriminating racial types, the interracial correlations are of more importance than the intraracial values dealt with above. Unfortunately the material collected is quite insufficient to give reliable estimates of the former. Four of the scatter diagrams given by the mean values are in Figs. 1—4, and the constants are in Tables VIII—X. In calculating the last, all means available based on 15 or more skulls were used in order to obtain a sufficient number of series, and it must be admitted that the sampling errors of the less reliable means are probably large enough to distort the resulting interracial constants. Owing to this fact, the standard deviations in Table VIII are probably larger than they should be. Nevertheless, those for the frontal and premaxillary indices are less than any intraracial standard deviations found (cf. Table V), and for the simotic and rhinal indices the interracial values are less than nearly all the intraracial. For all these indices interracial variability tends to be decidedly less than intraracial variability, and the same is true for the absolute measurements (cf. Tables IV and IX) and for nearly all of the more usual cranial measurements. The interracial correlations in Table VIII are probably somewhat too low, owing to the fact that several unreliable means were included in computing them. They are surprisingly different from the intraracial values in Table VI. All of the former class are decidedly higher than the corresponding coefficients of the latter class, except in the case of the premaxillary with the simotic and rhinal indices respectively for which the interracial and intraracial coefficients are not clearly differentiated. This demonstrates again the impossibility of deriving any knowledge of interracial correlation from a knowledge of intraracial constants only. The coefficients for absolute measurements in Tables VII and X can be used to demonstrate the same point. While all the values in Table VII are positive, there are several quite high negative values in Table X, and it is of interest to note that these are nearly all with the maxillary facial breadth (*GB*). There is a clear indication that *in the type* the frontal and nasal subtenses and the simotic breadth decrease with an increase in *GB*.

* The relative values of the correlation coefficients will depend partly on the racial homogeneity of the series. The Egyptian *E* series is known to be more homogeneous than most available series. Comparisons with the Javanese can only be made in the case of three male coefficients of variation, the values being:

	GB	*SS*	*SC*
Egyptian *E*	4·90 ± ·08 (877)	23·10 ± 1·64 (50)	16·84 ± 1·17 (50)
Javanese	4·67 ± ·17	35·65 ± 1·45	22·75 ± ·87

GB for Egyptian *E* is given by Pearson and Davin in *Biometrika*, Vol. xvi. 1924, pp. 328—363, and *SS* and *SC* by Ryley, Bell and Pearson in the same *Journal* for 1913, Vol. ix. pp. 391—445.

TABLE IX. *Interracial Means and Variabilities of Absolute Measurements for Series made up by Fifteen or more Crania.*

Sex		Male			Female		
Constants		Mean	Standard Deviation	Coefficient of Variation	Mean	Standard Deviation	Coefficient of Variation
Frontal ...	Subtense	17·58 ± ·104 (76)	1·338 ± ·073 (76)	7·61 ± ·419 (76)	16·19 ± ·134 (37)	1·213 ± ·095 (37)	7·49 ± ·591 (37)
	Chord ...	97·90 ± ·161 (76)	2·084 ± ·114 (76)	2·13 ± ·117 (76)	93·13 ± ·186 (37)	1·680 ± ·132 (37)	1·80 ± 1·412 (37)
Simotic ...	Subtense	3·55 ± ·070 (75)	·895 ± ·049 (75)	25·21 ± 1·474 (75)	2·85 ± ·095 (36)	·847 ± ·067 (36)	29·72 ± 2·563 (36)
	Chord ...	8·81 ± ·074 (75)	·949 ± ·052 (75)	10·77 ± ·600 (75)	8·73 ± ·129 (36)	1·147 ± ·091 (36)	7·61 ± ·608 (36)
Rhinal ...	Subtense	20·13 ± ·271 (45)	2·697 ± ·192 (45)	13·40 ± ·970 (45)	17·61 ± ·438 (14)	2·432 ± ·310 (14)	13·81 ± 1·794 (14)
	Chord ...	56·57 ± ·271 (45)	2·691 ± ·191 (45)	4·76 ± ·290 (45)	54·63 ± ·516 (14)	2·862 ± ·365 (14)	5·24 ± ·670 (14)
Premaxillary	Subtense	34·32 ± ·149 (70)	1·843 ± ·105 (70)	5·37 ± ·307 (70)	32·68 ± ·180 (33)	1·529 ± ·127 (33)	4·68 ± ·389 (33)
	Chord ...	96·20 ± ·246 (70)	3·048 ± ·174 (70)	3·17 ± ·181 (70)	91·57 ± ·333 (33)	2·833 ± ·235 (33)	3·09 ± ·257 (33)

TABLE X. *Interracial Correlations of Absolute Measurements for Series made up by Fifteen or more Crania.*

Characters	Sex	Sub. *IOW*	*SS*	Sub. *MOW*	Sub. *GB*	*IOW*	*SC*	*MOW*	*GB*
Sub. *IOW*	♂	—	+ ·744 ± ·035 (75)	+ ·690 ± ·053 (45)	+ ·156 ± ·079 (70)	+ ·124 ± ·076 (76)	+ ·462 ± ·061 (75)	− ·081 ± ·100 (45)	− ·592 ± ·052 (70)
	♀	—	+ ·550 ± ·078 (36)	—	− ·032 ± ·117 (33)	+ ·313 ± ·100 (37)	+ ·579 ± ·075 (36)	—	− ·517 ± ·086 (33)
SS	♂	+ ·744 ± ·035 (75)	—	+ ·862 ± ·026 (45)	− ·166 ± ·078 (70)	− ·049 ± ·078 (75)	+ ·578 ± ·052 (75)	− ·553 ± ·070 (45)	− ·481 ± ·070 (70)
	♀	+ ·550 ± ·078 (36)	—	—	− ·225 ± ·111 (33)	− ·170 ± ·109 (36)	+ ·602 ± ·072 (36)	—	− ·433 ± ·095 (33)
Sub. *MOW*	♂	+ ·690 ± ·053 (45)	+ ·862 ± ·026 (45)	—	− ·142 ± ·099 (45)	− ·016 ± ·101 (45)	+ ·220 ± ·096 (45)	− ·528 ± ·073 (45)	− ·547 ± ·070 (45)
	♀	—	—	—	—	—	—	—	—
Sub. *GB*	♂	+ ·156 ± ·079 (70)	− ·166 ± ·078 (70)	− ·142 ± ·099 (45)	—	+ ·441 ± ·065 (70)	− ·081 ± ·080 (70)	+ ·455 ± ·080 (45)	+ ·048 ± ·080 (70)
	♀	− ·032 ± ·117 (33)	− ·225 ± ·111 (33)	—	—	+ ·210 ± ·112 (33)	− ·140 ± ·115 (33)	—	+ ·168 ± ·114 (33)
IOW	♂	+ ·124 ± ·076 (76)	− ·049 ± ·078 (75)	− ·016 ± ·101 (45)	+ ·441 ± ·065 (70)	—	− ·111 ± ·077 (75)	+ ·296 ± ·092 (45)	+ ·189 ± ·078 (70)
	♀	+ ·313 ± ·100 (37)	− ·170 ± ·109 (36)	—	+ ·210 ± ·112 (33)	—	− ·027 ± ·112 (36)	—	+ ·119 ± ·116 (33)
SC	♂	+ ·462 ± ·061 (75)	+ ·578 ± ·052 (75)	+ ·220 ± ·096 (45)	− ·081 ± ·080 (70)	− ·111 ± ·077 (75)	—	+ ·127 ± ·099 (45)	− ·429 ± ·066 (70)
	♀	+ ·579 ± ·075 (36)	+ ·602 ± ·072 (36)	—	− ·140 ± ·115 (33)	− ·027 ± ·112 (36)	—	—	− ·667 ± ·065 (33)
MOW	♂	− ·081 ± ·100 (45)	− ·553 ± ·070 (45)	− ·528 ± ·073 (45)	+ ·455 ± ·080 (45)	+ ·296 ± ·092 (45)	+ ·127 ± ·099 (45)	—	+ ·089 ± ·100 (45)
	♀	—	—	—	—	—	—	—	—
GB	♂	− ·592 ± ·052 (70)	− ·481 ± ·070 (70)	− ·547 ± ·070 (45)	+ ·048 ± ·080 (70)	+ ·189 ± ·078 (70)	− ·429 ± ·066 (70)	+ ·089 ± ·100 (45)	—
	♀	− ·517 ± ·086 (33)	− ·433 ± ·095 (33)	—	+ ·168 ± ·114 (33)	+ ·119 ± ·116 (33)	− ·667 ± ·065 (33)	—	—

(7) *Racial Comparisons of the Measurements of Facial Flattening.* A description of the material measured is given in Section (3) above, and the mean measurements from which racial comparisons can be made are in Table I. Enough series are available, representing all the principal groups of modern races, to give a definite estimate of the value of the new criteria for anthropological purposes. Figs. 1—4 give scatter distributions of the mean indices taken in pairs, and no detailed descriptions of, or comments on, them need be made. But it is necessary to supplement the diagrams by discussing the order of differences which may be considered to indicate statistical differentiation in the case of each index.

Fig. 1 shows the interracial correlation of the rhinal and simotic indices. All series for which both means are based on 15 or more skulls are shown, but the majority of them are made up by considerably more than 15 individuals, the actual numbers being given in Table I. In order to obtain an estimate of differences which are statistically significant, it will be sufficient to consider the 11 male European series which are all in the top right-hand corner of the diagram. Taking the rhinal index first, it is found that out of the 55 possible differences between the means, 31 are less than 2·5 times their probable errors, 9 show values of the ratio of the difference to its probable error between 2·5 and 4·0, and for the remaining 15 the ratios are greater than 4·0. If we suppose that a difference is significant when it exceeds 2·5 times its probable error, then by the rhinal index the Dutch series is differentiated from all the others except the Russian; the Spitalfields and Mediaeval English from all except one another and the Farringdon Street English; while the Italian and Farringdon Street English series are also differentiated. Six of the series—viz. the Russian, Anglo-Saxon, Scottish, Irish, Romano-British and that of British Soldiers—cannot be distinguished from one another by their rhinal indices, and neither the Italian series, on the one hand, nor the Farringdon Street English, on the other, can be distinguished from the same group. Turning to the simotic index, it is found that 36 of the 55 differences are less than 2·5 times their probable errors; 9 show ratios between 2·5 and 4·0, and the remaining 10 show ratios greater than 4·0. Taking the limit again at 2·5, the Italian series is differentiated from all others except the Russian, the Spitalfields and the Mediaeval English; the Spitalfields is differentiated from all others except the Italian, Russian and Mediaeval English; and the Mediaeval English is differentiated from all others except the Italian, Russian, Spitalfields and Farringdon Street English and the Romano-British series. Eight of the 11 series—viz. the Dutch, Russian, Anglo-Saxon, Scottish, Irish, Romano-British, Farringdon Street English and that of the British Soldiers—thus show no significant differences between their mean simotic indices. If larger numbers of individuals were available, it is quite probable, of course, that some of the mean simotic indices for these racial types would be differentiated. For the material available, however, it is quite evident that statistically significant differences are found between points which are separated, in the case of either variate, by distances which are a small fraction of the total range shown by the same variate. It will be quite safe to assume, without further examination, that all the European series differ most

significantly from all the Oriental and African negro series in the case of both rhinal and simotic indices.

The frontal and premaxillary indices, used in Fig. 2, may be treated in the same way, and it will be sufficient to restrict attention to the group of five male Indian series at the top of the diagram. For these, only three of the possible ten comparisons of mean frontal indices show differences in excess of 2·5 times their probable errors, and the greatest ratio found is 3·0. The Singalese series shows a difference from that of Hindus from Madras which may be considered just significant, and the same is true for the Punjabi series from those of Hindus from North-East India and Madras respectively. More significant differences are found between the mean pre-maxillary indices, only three of the ratios being less than 2·5, five being between 2·5 and 4·0, and two being greater than 4·0. Both the means for Singalese and Hindus from North-East India are differentiated from all for the other three series. All the mean frontal indices of the Indian series will clearly be differentiated most markedly from all the means for the Oriental series, and quite markedly from nearly all the European means. The premaxillary index is less able to differentiate these three groups of races, but significant differences for it will be shown between means which are separated by quite a small fraction—less than one-tenth, say—of the range shown by all races in the world.

The fact that the last statement is true for all four indices indicates—since many of the series with which we are dealing are very small—that the indices are well-defined racial characters and their uses for purposes of classification may now be considered. The characters of this kind with which anthropologists are familiar may be divided somewhat arbitrarily into the following classes:

(*a*) Characters which appear to be of little use for any purposes of racial classification, although they may differ quite markedly from race to race.

(*b*) Characters which serve to differentiate and provide suggestive arrangements of racial types belonging to any single family of races, but which are practically valueless for the purpose of classifying the different families of races. Such are, notably, the cephalic index and stature, the ranges for these two being almost as great for European races, for example, as for all races in the world.

(*c*) Characters which are practically constant for races belonging to any single family of races, but which provide suggestive orders when the different families are compared. Such are, notably, skin colour, the nasal index and the nasal angle or other measurements of prognathism. Characters of this kind alone are capable of approximately grading all races in the world in order of their "primitiveness*."

Fig. 1 suggests forcibly that both the rhinal and simotic indices belong to the last of these classes. The interracial correlation between them for male means was

* See comments on this point by G. M. Morant in *Annals of Eugenics*, Vol. II. (1927), pp. 334—336, where it is concluded that the nasal index and nasal angle are the best fitted of the usual cranial characters to serve the purpose mentioned.

found to be +·868, and it is clear that the orders in which the two considered singly will arrange the races are very similar. African negro types have the most depressed nasal bones, then the Oriental races, which are remarkably similar to one another in this respect, follow and lead on to the American, Oceanic and Egyptian groups which are not distinguished from one another; the three Indian series all have more prominent nasal bones, and finally come the European types with the most prominent to be found in the world as far as our evidence can tell. This character of nasal prominence evidently tends to be fairly constant for races belonging to the same family—such as the European, Oriental or African negro groups—while it makes some very clear inter-group distinctions. Anyone unaccustomed to dealing with quantitative anthropological characters in this way may fail to appreciate the significance of the diagram we are considering. The vast majority of the measurements of size or shape usually provided fail entirely to supply any suggestive arrangements, when considered singly or in pairs, in the case of a sample drawn from all parts of the world. The cephalic index or stature, for example, will give an apparently meaningless jumble for such a sample of racial types. But the rhinal and simotic indices appear to be as capable of making inter-group distinctions as any other craniological characters known. The only other ones which can compare with them in this respect are the nasal index and the nasal angle or other measurements of prognathism. These last two are known to be highly correlated interracially with one another*, and it is probable that both are highly correlated with the rhinal and simotic indices. But the measurements of nasal prominence seem to give rather more suggestive orders than either the nasal index or angle†. The rhinal subtense was found to be rather an unsatisfactory measurement, since the tips of the nasal bones are so often broken, and it is sometimes difficult to decide whether they are defective or not. Since the resulting index has now also been shown to be highly correlated with the simotic index we do not recommend its further use.

The order in which the mean frontal indices arrange the racial types can be appreciated from Fig. 2. This is still a suggestive one, though it has evidently not the significance of those given by the rhinal and simotic indices. The Indian, European and Oriental groups are kept quite distinct and the African negroes occupy a restricted range between the last two. It is interesting to note that the Indian series all fall at one extreme of the distribution, while all belonging to the second Asiatic group—the Oriental—are at the other extreme. The primitive Oceanic and African negro series occupy intermediate positions and several of them are not differentiated by this character from European types. It is evident that the frontal index can give no indication whatever of "primitiveness." The position with regard to the premaxillary index is similar to that found in the case of most anthropological characters. It has been shown above that a small fraction of the total range may be taken to indicate a significant difference, so the index is as capable of differentiating races as most measurements. But it fails entirely to

* Morant found a coefficient of +·747 between the nasal index and nasal angle, *loc. cit.* p. 336.

† Cf. our Fig. 1 with the tables of nasal indices and angles given by Morant, *loc. cit.* pp. 335 and 336.

arrange them in an order which can be supposed to aid classification. No two of the principal families of races are distinguished and the range for African negroes is almost as great as for all races in the world. The premaxillary index is seen to be the least valuable of the four we are considering for purposes of inter-group comparisons and, like the others, it does not provide any interesting intra-group arrangements. It thus falls into class (*a*) defined above. What interest there is in the arrangement shown in Fig. 2 is given to it by the frontal index alone.

Fig. 3. shows the interracial correlation of the two most suggestive indices—the frontal and simotic. The coefficient between them for male types was found to be + ·699. Four of the principal families of races are seen to occupy discrete areas with somewhat of a jumble of Oceanic, American and Egyptian types falling between them. It may be doubted whether any other pair of craniological characters that have hitherto been investigated can effect a more interesting arrangement than this. The compactness of the points representing the Oriental series and their wide separation from the Indian group are particularly interesting facts. It has been shown by coefficients of racial likeness that when all the more important of the usual craniometric characters are considered together no sharp line of division can be drawn between these two Asiatic families of races, since the Nepalese type closely resembles that of Hindus of Bengal, on the one hand, and Tibetans of the *A* type on the other, while the last has close affinities with several Oriental types*. The relationships suggested by Fig. 3 alone are entirely different from these; the Tibetans *A* are as widely separated from the Indians as any other Oriental series, and, though the Nepalese point does diverge slightly in the direction of the Indians, it might well be considered a true member of the Oriental group. The frontal and simotic indices thus provide a means of distinguishing the two Asiatic families of races in the clearest possible manner, although a considerable number of other measurements fail to make a sharp division between them. This example warns us not to attempt to base any scheme of racial classification on the evidence of the two indices *alone*. They clearly provide valuable indications for the purpose in view—perhaps as valuable as those which can be given by any pair of characters known—but racial relationships can only be estimated by taking into account the evidence provided by a considerable number of characters which may not be highly correlated interracially with one another. It would be a gross mistake to assume that the New British are closely allied to any Oriental people merely because its point falls well inside the restricted area covered by all the Oriental races in Fig. 3, though mistakes of that kind are not uncommonly made.

Fig. 4 shows the interracial correlation of the premaxillary and simotic indices and it is obviously of far less interest than Fig. 3. The fact that the African Negro, Oriental and European groups are separated is seen to be due almost entirely to the separation made between them by the simotic indices. The premaxillary index may be supposed to give a measure of sub-nasal prognathism

* See T. L. Woo and G. M. Morant: "A Preliminary Classification of Asiatic Races Based on Cranial Measurements." *Biometrika*, Vol. xxiv. (1932), pp. 108—134.

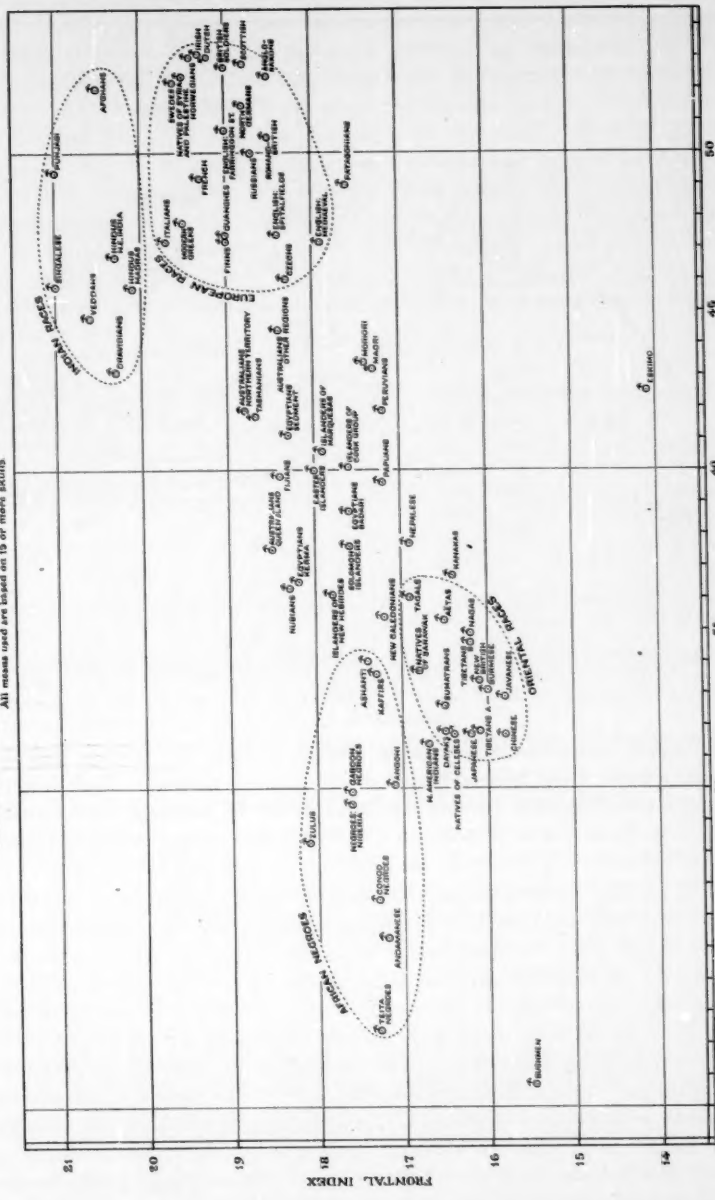

THE INTERRACIAL CORRELATION OF FRONTAL AND SIMOTIC INDICES.
All means used are based on 15 or more skulls
EUROPEAN RACES
AFRICAN NEGROES
ORIENTAL RACES
FRONTAL INDEX
SIMOTIC INDEX
20
25
30
35
40
45
50
21
20
19
18
17
16
15
14

Fig. 3.

THE INTERRACIAL CORRELATION OF PREMAXILLARY AND SIMOTIC INDICES.
All means used are based on 15 or more skulls.

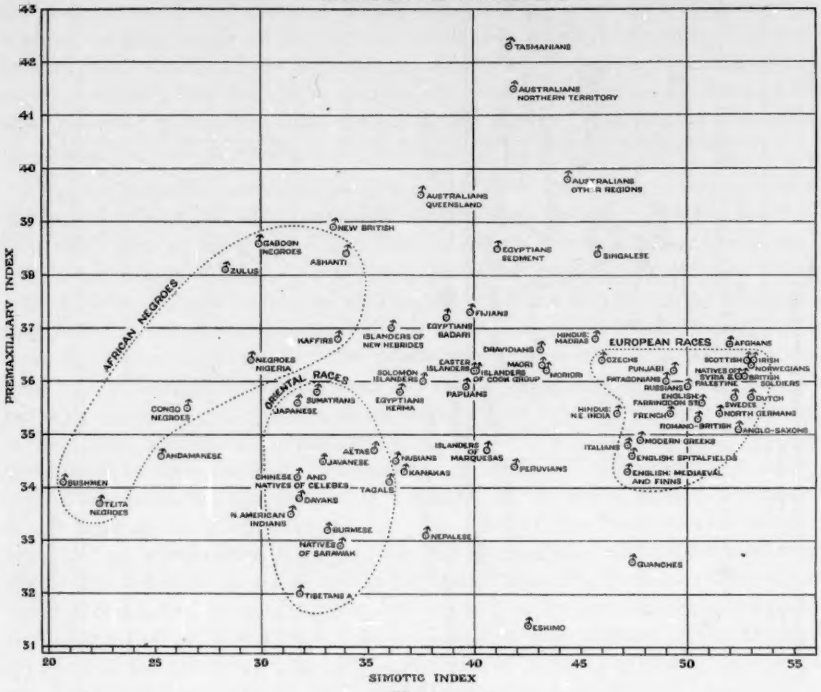


Fig. 4.

and the location of the Tasmanian and Australian series at the extreme end of the range would hence have been anticipated, but it is surprising to find that both Teita Negroes and Bushmen have lower means than all the European series available. This index does not appear to be of much value for purposes of racial classification and hence its further use cannot be advocated.

Finally, we may consider racial differences of the absolute measurements from which the four indices are derived. The rhinal subtense and chord may be omitted, since they can be given for considerably fewer series than can the other measurements, and the premaxillary subtense is not of special interest. The breadth (*GB*) associated with that subtense is included, however, since it is a measurement which has been given for far more series than the new ones taken for the purpose of this paper. The ranges for different groups of races given in Table XI are found from the data given in Table I. The frontal subtense (Sub. *IOW*), for the data available, is seen to make an absolute distinction—i.e. to show ranges which do not overlap—between the African Negro and Oriental, African Negro and Indian, Oriental and European, Oriental and Indian, and Oceanic and Indian groups, while precisely the same distinctions and no others are made by the mean frontal indices (see

Fig. 2). But the frontal breadth (*IOW*) is only able to distinguish the African Negro and Indian groups. The simotic subtense makes absolute distinctions between the African Negro and European, African Negro and Indian, Oriental and European, Oriental and Indian, and Oceanic and European groups, while the mean simotic indices make all these distinctions and also one between the African Negro and Oceanic groups as well (see Fig. 3). But the simotic chord is not able to distinguish any of the groups. These rough comparisons suggest forcibly that the frontal and simotic indices are valuable racial criteria by virtue of differences between the subtenses, but not between the chords involved. The case is different for the premaxillary characters, which were not valuable racial criteria, since the index makes no distinctions between any of the groups (see Fig. 2), but the breadth (*GB*) differentiates the Oriental group from both the European and Indian (see Table XI).

TABLE XI.

Ranges of Male Mean Absolute Measurements (all based on Fifteen or more Crania for different Groups of Races.*

Racial Groups	Sub. *IOW*	*SS*	*IOW*	*SC*	*GB*
African Negro†	17·1—18·6 (8)	2·2—2·9 (8)	98·6—103·0 (8)	8·5— 9·5 (8)	93·9— 98·5 (7)
Oriental‡ ...	15·1—16·5 (14)	2·4—3·4 (14)	94·3— 99·5 (14)	7·6— 9·6 (14)	97·4—101·5 (12)
Oceanic ...	16·0—18·8 (16)	2·7—4·1 (16)	97·3—101·8 (16)	6·5— 9·3 (16)	89·8—102·6 (15)
European ...	17·7—19·3 (19)	4·5—4·9 (18)	96·6—100·0 (19)	8·6—10·3 (18)	91·7— 97·2 (18)
Indian§ ...	18·9—20·2 (7)	3·7—5·0 (7)	93·2— 97·6 (7)	8·5— 9·8 (7)	93·0— 96·9 (6)

* The American series were omitted since very few are available, as are some others, such as the Egyptian series, which cannot be placed in any of the groups included. The numbers in brackets are the numbers of series from which the extremes were deduced.

† Excluding the Bushman series. ‡ Excluding the Andamanese but including the Celebes series.

§ Including the Afghan series.

8. *Conclusions.* The transverse "flatness" of the facial skeleton has long been recognised to be a valuable criterion for the purpose of aiding the discrimination and classification of the races of modern man. This research was undertaken with the object of examining it by quantitative methods on an adequate number of crania. Four pairs of absolute measurements were taken, whenever possible, on each specimen giving four indices which express a "median" subtense as a percentage of a corresponding transverse breadth. One of these indices—the simotic—had been used previously and the others are termed the frontal, rhinal and premaxillary indices. All the subtenses were found with a pair of co-ordinate calipers made for the purpose. The material measured (preserved in various English and one Dutch museum) consists of 4266 male and 1630 female specimens divided into 83 series which represent all the principal groups of races of modern man. An analysis of the statistical constants leads to the following conclusions:

(i) Five of the eight absolute measurements show sex ratios of the usual order, but for the simotic chord the ratios are small and for the simotic and rhinal sub-

tenses they are particularly large. The simotic and rhinal indices hence show marked sexual differences between means, while the premaxillary index shows slight sexual differentiation and the frontal none at all. The nasal bones are, on the average, more protruding in the male than in the female of the same race, but the premaxillary region tends to be less protruding in the male.

(ii) In the case of the indices male variation shows a distinct tendency to be greater than female, but for absolute measurements—judging by coefficients of variation—female variation is either greater than male, or else there is sexual equality in variability.

(iii) There is no evidence that the variabilities of the characters examined differentiate "primitive" and "advanced" races.

(iv) The intraracial correlations of the absolute measurements and indices are decidedly lower than would have been anticipated.

(v) The interracial variabilities tend to be decidedly smaller than the intraracial variabilities.

(vi) The interracial correlations of the indices are clearly differentiated from, and they tend to be decidedly higher than, the corresponding intraracial correlations. The correlations of the two classes in the case of the absolute measurements show many marked differences.

(vii) All four indices show significant differences between the majority of the means that are available, so they are well-marked racial characters.

(viii) While showing no marked differences between races belonging to the same family, or group, of races, the rhinal and simotic indices—which are highly correlated interracially—appear as capable of making inter-group distinctions as any other craniological characters that have been examined (see Fig. 1).

(ix) The frontal index also shows small intra-group differences, and it makes several clear inter-group distinctions, though these are not so well-defined as those given by the rhinal and simotic indices (see Fig. 2).

(x) The premaxillary index fails to make distinctions between the different groups of races and it does not provide any intra-group arrangements to which significance can be attached (see Fig. 2).

(xi) There would be little advantage in determining the rhinal and premaxillary indices for larger numbers of skulls, but the simotic and frontal indices are valuable characters which should be recorded in the routine descriptions of racial series of crania. The arrangement provided by the last two (see Fig. 3) is as interesting as that given by any other pair of cranial measurements. Some distinctions are made by the simotic and frontal indices which are not effected by any other cranial characters that have been investigated.

(xii) It is shown that the simotic and frontal indices are valuable racial criteria by virtue of racial differences between the subtenses—but not between the chords—from which they are formed.

MISCELLANEA.

(i) On certain Measures of Dependence between Statistical Variables.

By J. F. STEFFENSEN, Copenhagen.

1. In a lecture delivered at "Institut Henri-Poincaré" and published in the *Annales* of that Institute (1933, pp. 319—331) I have suggested a modification of Karl Pearson's Mean Square Contingency. The object of the modification was to present the measure in such a form that it remains applicable also when the number of distinct values which the statistical variables can assume becomes infinite. It has been pointed out to me recently by Mrs Pollaczek that the measure in the form I had proposed does not always, as I had somewhat incautiously asserted, assume the value 1 in the case of complete dependence between the variables. A slight modification suffices, however, to meet this objection, and the measure is thus reduced to its original form which I communicated to Professor Cramér at the beginning of 1930 without publishing it. In the present Note I intend to establish the measure from first principles in its correct form, and thereafter to propose a new measure which presents certain advantages over the other.

2. With a view to comparisons with the lecture, it is preferable to retain the same notation, and as this differs somewhat from notations employed by other authors, I shall begin by explaining the symbols preferred.

Let there be a compound event depending on two statistical variables x and y of which x must assume one of the distinct values

$$x_1, x_2, x_3, \ldots,$$

and y one of the distinct values $y_1, y_2, y_3, \ldots,$

each with a certain non-vanishing probability, the sum of the probabilities being unity for each sequence of values. The number of values may be finite or not, but it is assumed to exceed one in each sequence, as otherwise x or y would not be a statistical variable.

The probability that x assumes the value x_i and y the value y_j is denoted by p_{ij}. The probability that x assumes the value x_i irrespective of what value y assumes is obtained by summing p_{ij} with respect to all j. A summation of this nature is denoted by replacing the letter with respect to which the summation is performed by an asterisk. We thus have

$$p_{i*}=\sum_j p_{ij} \quad \text{and similarly} \quad p_{*j}=\sum_i p_{ij}.$$

We also note

$$p_{**}=\sum_{ij} p_{ij}=\sum_i p_{i*}=\sum_j p_{*j}=1.$$

It was assumed that p_{i*} and p_{*j} cannot vanish; they are also smaller than unity, because there are more than one possible values for each of the variables. We have, therefore,

$$0<p_{i*}<1, \quad 0<p_{*j}<1 \qquad \ldots\ldots(1),$$

for all values of i and j.

The probability that y assumes the value y_j *when it is known* that x assumes the value x_i is denoted by $p_{(i)j}$. We thus have

$$p_{ij}=p_{i*}p_{(i)j} \qquad \ldots\ldots(2),$$

and similarly

$$p_{ij}=p_{*j}p_{i(j)} \qquad \ldots\ldots(3).$$

The extreme case where x and y are *completely independent* of each other is characterised by the relation

$$p_{ij}=p_{i*}p_{*j} \quad \text{.....(4),}$$

valid for all i and j.

The other extreme case occurs if one of the statistical variables is a function of the other. In this case, each value of x is always associated with the same value of y, and we may therefore, without loss of generality, associate x_i with y_i. The condition of *complete dependence* may then be written

$$p_{ii}=p_{i*}=p_{*i}; \quad p_{ij}=0 \text{ for } i\neq j \quad \text{.....(5).}$$

3. After these preliminaries, we may proceed to the measure. It is denoted by ψ and defined by the relation

$$\psi^2=\sum_{ij} p_{ij}\phi_{ij}^2 \quad \text{.....(6),}$$

where

$$\phi_{ij}^2=\frac{(p_{ij}-p_{i*}p_{*j})^2}{p_{i*}(1-p_{i*})p_{*j}(1-p_{*j})} \quad \text{.....(7).}$$

We propose to show that ψ^2 possesses the following properties:

(I) ψ^2 is always comprised between 0 and 1.

(II) ψ^2 vanishes if the variables are completely independent and only in that case.

(III) ψ^2 assumes the value 1 in the case of complete dependence, and only in that case.

As regards the property (I), we observe that ψ^2 is an arithmetical mean, with positive or zero coefficients, of the quantities ϕ_{ij}^2. It therefore suffices to prove that $\phi_{ij}^2\leqq 1$ for all values of i and j.

Now, in the cases where $p_{ij}\geqq p_{i*}p_{*j}$, we write ϕ_{ij}^2 in the form

$$\phi_{ij}^2=\frac{p_{ij}-p_{i*}p_{*j}}{p_{i*}(1-p_{*j})}\cdot\frac{p_{ij}-p_{i*}p_{*j}}{p_{*j}(1-p_{i*})} \quad \text{.....(8).}$$

As $p_{ij}\leqq p_{i*}$, we have

$$p_{ij}-p_{i*}p_{*j}\leqq p_{i*}(1-p_{*j}),$$

so that the first factor in (8) cannot exceed unity.

Further, as $p_{ij}\leqq p_{*j}$, we have

$$p_{ij}-p_{i*}p_{*j}\leqq p_{*j}(1-p_{i*}),$$

so that the second factor in (8) cannot exceed unity. We therefore have $\phi_{ij}^2\leqq 1$, if $p_{ij}\geqq p_{i*}p_{*j}$.

Considering, next, the cases where $p_{ij}\leqq p_{i*}p_{*j}$, we write ϕ_{ij}^2 in the form

$$\phi_{ij}^2=\frac{p_{i*}p_{*j}-p_{ij}}{p_{i*}p_{*j}}\cdot\frac{p_{i*}p_{*j}-p_{ij}}{(1-p_{i*})(1-p_{*j})} \quad \text{.....(9).}$$

Here, the first factor obviously cannot exceed unity. As regards the second factor, we begin by observing that

$$p_{i*}-p_{ij}\leqq 1-p_{*j},$$

because the right-hand side is obtained by summation of the left-hand side with respect to i. We have therefore

$$p_{i*}p_{*j}-p_{ij}\leqq 1-p_{i*}-p_{*j}+p_{i*}p_{*j}$$

or

$$p_{i*}p_{*j}-p_{ij}\leqq (1-p_{i*})(1-p_{*j}),$$

so that the second factor in (9) cannot exceed unity.

We therefore have $\phi_{ij}^2\leqq 1$ also if $p_{ij}\leqq p_{i*}p_{*j}$; and hence, in all cases.

Proceeding to the property (II), it is obvious that ψ^2 vanishes if $p_{ij}=p_{i*}p_{*j}$ for all values of i and j, that is, if the variables are completely independent.

In order to prove the converse proposition, we assume that ψ^2 vanishes, and have to show that then $p_{ij}=p_{i*}p_{*j}$ for all values of i and j. We begin by proving that if $\psi^2=0$, no p_{ij} can vanish*.

Let us for a moment assume that, for instance, $p_{rs}=0$; this assumption leads to a contradiction if $\psi^2=0$. We have, in fact,

$$\sum_{ij}(p_{ij}-p_{i*}p_{*j})=0,$$

and as the term $(p_{rs}-p_{r*}p_{*s})$

is negative, because $p_{rs}=0$, there must be at least one positive term in the sum. Let, for instance, the term

$$(p_{nm}-p_{n*}p_{*m})$$

be positive. In that case ϕ_{nm}^2 does not vanish, and ψ^2 contains therefore the positive term $p_{nm}\phi_{nm}^2$. Therefore $\psi^2>0$; but we had assumed $\psi^2=0$, and we have thus arrived at a contradiction.

As no p_{ij} can vanish, if $\psi^2=0$, it follows from (6) that all ϕ_{ij}^2 vanish if $\psi^2=0$, that is, we have $p_{ij}=p_{i*}p_{*j}$ for all values of i and j.

Turning, finally, our attention to the property (III), we first have to prove that if the dependence is complete, then $\psi^2=1$. Now, making use of (5), (6) reduces to

$$\sum_i p_{ii}\frac{(p_{ii}-p_{ii}^2)^2}{p_{ii}(1-p_{ii})\,p_{ii}(1-p_{ii})}=\sum_i p_{ii}=1.$$

Conversely, let us assume that $\psi^2=1$, that is

$$\sum_{ij}p_{ij}\phi_{ij}^2=1.$$

As $\sum_{ij}p_{ij}=1$, and as $\phi_{ij}^2\leqq 1$, this relation cannot hold, unless $\phi_{ij}^2=1$ for all values of i and j for which p_{ij} does not vanish. Now, if $p_{ij}>0$, we cannot at the same time have $p_{ij}\leqq p_{i*}p_{*j}$ and $\phi_{ij}^2=1$, as the first factor in (9) is less than unity. We therefore have $p_{ij}>p_{i*}p_{*j}$, and as each of the factors in (8) must equal unity, if $\phi_{ij}^2=1$, we have $p_{ij}=p_{i*}=p_{*j}$, that is, y_j occurs certainly, if x_i occurs.

These considerations show that if $\psi^2=1$, then there is complete dependence between the variables.

4. In order to approximate to ψ by means of a given experience, we replace the probabilities by the corresponding relative frequencies. Let H_{ij} be the absolute frequency of the combination (x_i, y_j), and let us write

$$H_{i*}=\sum_j H_{ij},\quad H_{*j}=\sum_i H_{ij},\quad N=H_{**},$$

so that N is the total number of observations. If Ψ denotes the approximation to ψ obtained in this way, we find by (6) and (7)

$$\Psi^2=\frac{1}{N}\sum_{ij}\frac{H_{ij}(NH_{ij}-H_{i*}H_{*j})^2}{H_{i*}(N-H_{i*})\,H_{*j}(N-H_{*j})}\qquad\text{.....(10).}$$

For the reasons explained in the lecture, no attempt has been made to estimate the mean error of this expression.

5. The calculation of Ψ by (10), although not prohibitive, is more laborious than desirable, and I have therefore endeavoured to establish another measure which leads to simpler calculations. The new measure which I shall denote by ω is defined as

$$\omega=\frac{\sum|p_{ij}-p_{i*}p_{*j}|}{\sum(p_{ij}-p_{ij}^2)+\sum p_{i*}p_{*j}}\qquad\text{.....(11),}$$

* I owe this observation to Mr N. P. Bertelsen with whom I have discussed various points in this paper.

where Σ denotes summation with respect to all i and j, $\bar{\Sigma}$ summation with respect to all i and j for which $p_{ij} > p_{i*}p_{*j}$, and $\underline{\Sigma}$ summation with respect to the remaining i and j, that is, the values for which $p_{ij} \leqq p_{i*}p_{*j}$.

It can be proved that ω possesses the same three fundamental properties (I), (II) and (III) as ψ. Before proceeding to the proof, it is useful to point out that ω may be written in various other forms. For the sake of argumentation it is preferable to write

$$\omega = \frac{\bar{\Sigma}(p_{ij} - p_{i*}p_{*j}) + \underline{\Sigma}(p_{i*}p_{*j} - p_{ij})}{\bar{\Sigma}(p_{ij} - p_{ij}^2) + \underline{\Sigma} p_{i*}p_{*j}} \quad \text{.....(12).}$$

Further, as

$$\Sigma(p_{ij} - p_{i*}p_{*j}) = 0,$$

the aggregate of positive terms in this sum must cancel the aggregate of negative ones; hence (11) may be written

$$\omega = 2\frac{\bar{\Sigma}(p_{ij} - p_{i*}p_{*j})}{\bar{\Sigma}(p_{ij} - p_{ij}^2) + \underline{\Sigma} p_{i*}p_{*j}} \quad \text{.....(13).}$$

If, in this expression, we eliminate $\underline{\Sigma}$ by the relation $\Sigma = \bar{\Sigma} + \underline{\Sigma}$ and observe that $\Sigma p_{i*}p_{*j} = 1$, we obtain

$$\omega = 2\frac{\bar{\Sigma}(p_{ij} - p_{i*}p_{*j})}{\bar{\Sigma}(p_{ij} - p_{i*}p_{*j}) + 1 - \bar{\Sigma} p_{ij}^2} \quad \text{.....(14).}$$

The convergence of all the sums is obvious, as

$$\Sigma|p_{ij} - p_{i*}p_{*j}| < \Sigma p_{ij} + \Sigma p_{i*}p_{*j} = 2,$$

$$\bar{\Sigma}(p_{ij} - p_{ij}^2) + \underline{\Sigma} p_{i*}p_{*j} < \Sigma p_{ij} + \Sigma p_{i*}p_{*j} = 2,$$

and

$$\bar{\Sigma} p_{ij}^2 < \Sigma p_{ij} = 1.$$

As regards the proof of the first property, or (I), we observe that $p_{ij} \leqq p_{i*}$, $p_{ij} \leqq p_{*j}$, so that $p_{ij}^2 \leqq p_{i*}p_{*j}$. It follows that each term in the nominator of (12) cannot exceed the corresponding term in the denominator, so that $0 \leqq \omega \leqq 1$.

The second property (II) is obvious, as (11) shows that ω vanishes if $p_{ij} = p_{i*}p_{*j}$ for all values of i and j, and only then.

In order to prove the third property (III), we must first show that $\omega = 1$ if there is complete dependence between x and y. Now in that case we have, according to (5), $p_{ij} = 0$ if $i \neq j$, and the corresponding terms which belong to $\underline{\Sigma}$ in (12) become equal in nominator and denominator. If, on the other hand, $i = j$, we have $p_{ii} = p_{i*} = p_{*i}$, and the corresponding terms which belong to $\bar{\Sigma}$ in (12) become equal to $(p_{ii} - p_{ii}^2)$ both in nominator and denominator of (12). Hence $\omega = 1$, if the dependence is complete.

We finally have to prove that if $\omega = 1$, then the dependence is complete. But, observing that all the terms in the sums in (12) are positive or zero, and that each term in the nominator cannot exceed the corresponding term in the denominator, it becomes clear that, if $\omega = 1$, then each term in the nominator must be equal to the corresponding term in the denominator. Hence, we have *either* $p_{ij} = 0$ *or else* $p_{ij}^2 = p_{i*}p_{*j}$; but on multiplying (2) and (3) with each other it is seen that we can only have $p_{ij}^2 = p_{i*}p_{*j}$ if $p_{i(j)} = 1$, $p_{i(j)} = 1$, so that

$$p_{ij} = p_{i*} = p_{*j}.$$

The dependence is, therefore, complete.

6. In order to approximate to ω by means of a given experience, we replace, as in paragraph 4, the probabilities by the corresponding relative frequencies. The simplest formula to employ for the numerical work is (14). If Ω denotes the approximation to ω obtained in this way, we find

$$\Omega = 2\frac{\bar{\Sigma} N H_{ij} - \bar{\Sigma} H_{i*}H_{*j}}{\bar{\Sigma} N H_{ij} - \bar{\Sigma} H_{i*}H_{*j} + N^2 - \bar{\Sigma} H_{ij}^2} \quad \text{.....(15),}$$

where $\bar{\Sigma}$ means summation with respect to all values of i and j for which $NH_{ij} > H_{i*}H_{*j}$.

It is seen that the calculation of Ω by (15) is considerably easier than the calculation of Ψ by (10).

7. In order to apply the measures ψ and ω to continuous probabilities, we choose a constant interval h for x, and a corresponding interval k for y, and put

$$p_{ij}=\int_{jk}^{jk+k}\int_{ih}^{ih+h} f(x,y)\,dx\,dy,$$

$$p_{i*}=\int_{ih}^{ih+h} f(x,*)\,dx, \qquad p_{*j}=\int_{jk}^{jk+k} f(*,y)\,dy,$$

where the asterisk means integration over the whole range. Inserting these values in (6) and (11), it is seen that the values of ψ and ω depend on the values chosen for h and k. This dependence is very pronounced in the case of ψ, less so in the case of ω, so that ω is also in this respect preferable to ψ.

An entirely satisfactory measure in the case of continuous probabilities has not yet been found, as far as I am aware. The measures proposed by H. Cramér* and by Mrs Pollaczek† are subject to inconveniences to which these authors themselves draw attention.

(ii) Remarks on Professor Steffensen's Measure of Contingency.

EDITORIAL.

Let us consider a bivariate contingency table in which m_{ij} individuals are in the ij internal cell, and for which $m_{i.}$ individuals fall in the ith cell of one marginal total and $m_{.j}$ in the jth cell of the other marginal total.

Let $$M=S_{ij}(m_{ij})=S_i(m_{i.})=S_j(m_{.j}).$$

Consider the expression

$$\psi^2=S_{ij}\frac{m_{ij}\left(m_{ij}-\dfrac{m_{i.}m_{.j}}{M}\right)^2}{Mm_{i.}m_{.j}\left(1-\dfrac{m_{i.}}{M}\right)\left(1-\dfrac{m_{ij}}{M}\right)} \qquad \text{.....(i).}$$

Then from (i) it is possible to show that, if

(*a*) $m_{ij}=m_{i.}m_{.j}/M$ for all values of i and j, then the two variates are independent and $\psi^2=0$, and conversely if $\psi^2=0$, the variates are independent;

(*b*) $m_{ij}=m_{i.}=m_{.j}$, the two variates are such that if a value of one occurs in the ith marginal cell, the corresponding value of the second will always fall into the jth marginal cell, and the value of ψ^2 will be 1, and conversely if the value of ψ^2 be unity, then the above relation holds;

(*c*) the value of ψ^2 always lies between 0 and 1.

Now let us write $p_{ij}=m_{ij}/M$ and $p_{i.}=m_{i.}/M$ and $p_{.j}=m_{.j}/M$, then (i) becomes

$$\psi^2=S_{ij}\frac{p_{ij}(p_{ij}-p_{i.}p_{.j})^2}{p_{i.}p_{.j}(1-p_{i.})(1-p_{.j})} \qquad \text{.....(ii).}$$

(ii) is clearly in form identical with Professor Steffensen's ψ^2 value.

We may define p_{ij} as the chance of drawing an individual out of the ijth cell and $p_{i.}$ and $p_{.j}$ as the chances of drawing individuals from ith and jth marginal cells, when an individual must be returned, before a second drawing, to the cell from which it has been extracted.

* *Skandinavisk Aktuarietidskrift*, Vol. VII. (1924), p. 231.

† *Zeitschrift für angewandte Mathematik*, Vol. 13 (1933), p. 122.

Now there seems in this method of approaching the subject no restriction on the number of cells or on the smallness or largeness of the p's, or it seems allowable to pass to differentials.

Suppose then we have a frequency surface of x, y variates, where

$$z_1 = Mf_1(x, y),$$

and for the distributions of x and y alone, i.e. marginal total curves,

$$z_2 = Mf_2(x), \quad z_3 = Mf_3(y).$$

We can take

$$m_{ij} = Mf_1(x, y)\,dx\,dy,$$

$$m_{i.} = Mf_2(x)\,dx, \quad m_{.j} = Mf_3(y)\,dy,$$

and accordingly

$$\psi^2 = S_{ij} \frac{dx^3 dy^3 f_1(x, y)(f_1(x, y) - f_2(x) f_3(y))^2}{f_2(x) f_3(y)\,dx\,dy\,(1 - f_2(x)\,dx)(1 - f_3(y)\,dy)}$$

$$= \iint dx^2 dy^2 \left\{ \frac{f_1(x, y)(f_1(x, y) - f_2(x) f_3(y))^2}{f_1(x) f_3(y)} \right\} \text{ in the limit.}$$

The expression in curled brackets will be finite and accordingly the integral, owing to the presence of $dx^2 dy^2$ instead of $dx\,dy$, will vanish, however close or loose the bond between x and y.

The mean square contingency as I define it is given by

$$\phi^2 = S_{ij} \frac{\left(m_{ij} - \frac{m_{i.} m_{.j}}{M}\right)^2}{M \frac{m_{i.} m_{.j}}{M}}$$

$$= S_{ij} \frac{(p_{ij} - p_{i.} p_{.j})^2}{p_{i.} p_{.j}}$$

$$= \iint dx\,dy \frac{(f_1(x, y) - f_2(x) f_3(y))^2}{f_2(x) f_3(y)} \text{ in the limit,}$$

and remains finite when we pass to the limit. My coefficient of mean square contingency, C_2, is given by

$$C_2 = \sqrt{\frac{\phi^2}{1 + \phi^2}},$$

and this always lies between 0 and $\sqrt{1 - \frac{1}{n}}$, where n is the number of cells occupied if there be only one occupied cell in each row and in each column of the table. Only when n is indefinitely large, does C_2 range from 0 to unity, for example in cases of continuous distribution. So far from considering this a disadvantage of C_2, I hold it an essential property of a contingency coefficient; for as long as the cells are not indefinitely small, it is possible to arrange the individuals in them so as to increase the contingency. I cannot therefore hold that in choosing ψ^2 in the form (i) instead of C_2, we should gain any advantage from the fact that its upper limit is unity. I do not think a good coefficient of contingency ought to be unity until we introduce the conception of continuity, namely every y having attached to it a definite x, so that y is a continuous function of x. The fact that Professor Steffensen's ψ^2 is necessarily always zero when we pass to continuity, suggests that it will take a very low value in cases where we should expect a high value of the variate relationship. Indeed, the presence of the factor m_{ij} in the numerator of (i), or of p_{ij} in Professor Steffensen's expression, means there can be no contribution to ψ^2 from the ijth cell when $m_{i.}$, $m_{.j}$, or $p_{i.}$, $p_{.j}$ are finite but p_{ij} is zero. This seems a serious disadvantage in practical working, for if $m_{i.}$ and $m_{.j}$ are considerable, and yet m_{ij} zero, we should expect a considerable addition to the contingency, which is not provided by the Steffensen ψ^2.

Illustrations.

(1) Consider the contingency Table I. ψ^2 based on the 28 *occupied* cells is equal to ·12833. ϕ^2 based on the 100 cells which *might* be occupied with these marginal probabilities is equal to 2·983,333 and the Coefficient of Mean Square Contingency $C_2=\sqrt{\frac{\phi^2}{1+\phi^2}}=$·86542.

TABLE I.

·	·	·	·	·	·	·	·	·	1	2	3
·	·	·	·	·	·	·	1	2	2		5
·	·	·	·	·	·	1	2	1	·		4
·	·	·	·	·	1	2	1	·	·		4
·	·	·	·	1	2	1	·	·	·		4
·	·	·	1	2	1	·	·	·	·		4
·	·	1	2	1	·	·	·	·	·		4
·	1	2	1	·	·	·	·	·	·		4
2	2	1	·	·	·	·	·	·	·		5
2	1	·	·	·	·	·	·	·	·		3
4	4	4	4	4	4	4	4	4	4		40

Surely a high degree of relationship might be anticipated from such a table; but ψ^2 gives it a comparatively low place.

(2) Now let us introduce more rows of 1, 2, 1 so that instead of 10 fours in the horizontal marginal total we have 15. Table II shows the distribution. Here $\psi^2=$·138,003 and $\phi^2=$4·858,333, giving a Coefficient of Mean Square Contingency $C_2=\sqrt{\phi^2/(1+\phi^2)}=$·91066. Thus both ψ^2 and C_2 have risen, but ψ^2 remains low and C_2 is very high. The latter is in accordance with what we should expect, if this were a correlation table for continuous variates.

TABLE II.

·	·	·	·	·	·	·	·	·	·	·	·	1	2		3
·	·	·	·	·	·	·	·	·	·	1	2	2			5
·	·	·	·	·	·	·	·	·	1	2	1	·			4
·	·	·	·	·	·	·	·	1	2	1	·	·			4
·	·	·	·	·	·	·	1	2	1	·	·	·			4
·	·	·	·	·	·	1	2	1	·	·	·	·			4
·	·	·	·	·	1	2	1	·	·	·	·	·			4
·	·	·	·	1	2	1	·	·	·	·	·	·			4
·	·	·	1	2	1	·	·	·	·	·	·	·			4
·	·	1	2	1	·	·	·	·	·	·	·	·			4
·	1	2	1	·	·	·	·	·	·	·	·	·			4
2	2	1	·	·	·	·	·	·	·	·	·	·			5
2	1	·	·	·	·	·	·	·	·	·	·	·			3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	60

(3) We will now test what happens if we increase the contents of the central cell in the rows. Table III gives such a distribution. Here Steffensen's ψ^2 rises with a leap to ·500,940 while my Coefficient of Mean Square Contingency $C_2=$·9374. This result is not very informative. It indicates that ψ^2 is very sensitive to increase in the central probabilities, which we pushed up from $\frac{1}{30}$ to $\frac{1}{15}$, and to decrease in the other probabilities which we have reduced from $\frac{1}{60}$ to $\frac{1}{120}$.

TABLE III.

	·	·	·	·	·	·	·	·	·	·	·	1	1
	·	·	·	·	·	·	·	·	·	·	1	8	9
	·	·	·	·	·	·	·	·	·	1	8	1	10
	·	·	·	·	·	·	·	·	1	8	1	·	10
	·	·	·	·	·	·	·	1	8	1	·	·	10
	·	·	·	·	·	·	1	8	1	·	·	·	10
	·	·	·	·	·	1	8	1	·	·	·	·	10
	·	·	·	·	1	8	1	·	·	·	·	·	10
	·	·	·	1	8	1	·	·	·	·	·	·	10
	·	·	1	8	1	·	·	·	·	·	·	·	10
	·	1	8	1	·	·	·	·	·	·	·	·	10
	1	8	1	·	·	·	·	·	·	·	·	·	10
	8	1	·	·	·	·	·	·	·	·	·	·	9
	1	·	·	·	·	·	·	·	·	·	·	·	1
	10	10	10	10	10	10	10	10	10	10	10	10	120

But it does not in itself render obvious whether a rise from ·1380 to ·5009 is a more reasonable change than one from ·9107 to ·9374.

We will consider these changes on another series of tables.

(4) Consider Table IV. Here *à priori* I should anticipate a very considerable degree of relationship. Steffensen's $\psi^2 = ·20139$. The Mean Square Contingency, ϕ^2, $=2$ and $C_2 = ·81650$. Thus the ψ^2 is small and the C_2 very considerable.

TABLE IV.

	·	·	·	·	·5	·5
	·	·	·	·5	·5	1
	·	·	·5	·5	·	1
	·	·5	·5	·	·	1
	·5	·5	·	·	·	1
	·5	·	·	·	·	·5
	1	1	1	1	1	5

(5) We will now increase the number of rows of ·5, ·5 until we have 99 of them with a ·5 as the first and last row. Accordingly there will be 101 rows and 100 columns. Again I should anticipate that such a table must make a high degree of relationship. $\psi^2 = ·24750$, while $C_2 = ·98094$. My opinion is that ψ^2 gives a far too low measure of the degree of relationship of the two variates, and its values are not well distributed throughout the range 0 to 1.

TABLE V.

·	·	·	·	·	·	·	·	·	·5	·5
·	·	·	·	·	·	·	·	·5	·5	1
·	·	·	·	·	·	·	·5	·5	·	1
				·5, ·5 in 99 rows						·
				altogether						·
·	·5	·5	·	·	·	·	·	·	·	1
·5	·5	·	·	·	·	·	·	·	·	1
·5	·	·	·	·	·	·	·	·	·	·5
1	1	1	·	·	·	1	1	1		100

(6) Let us now see what happens when we increase the scatter. We will take four ·5's to a row. Table VI suggests a high degree of relationship, if it were based on a continuous variation.

TABLE VI.

	·	·	·	·	·	·	·	·	·	·5	.5
	·	·	·	·	·	·	·	·	·5	·5	1
	·	·	·	·	·	·	·	·5	·5	·5	1·5
	·	·	·	·	·	·	·5	·5	·5	·5	2
	·	·	·	·	·	·5	·5	·5	·5	·	2
	·	·	·	·	·5	·5	·5	·5	·	·	2
	·	·	·	·5	·5	·5	·5	·	·	·	2
	·	·	·5	·5	·5	·5	·	·	·	·	2
	·	·5	·5	·5	·5	·	·	·	·	·	2
	·5	·5	·5	·5	·	·	·	·	·	·	2
	·5	·5	·5	·	·	·	·	·	·	·	1·5
	·5	·5	·	·	·	·	·	·	·	·	1
	·5	·	·	·	·	·	·	·	·	·	·5
	2	2	2	2	2	2	2	2	2	2	20

Steffensen's ψ^2 sinks to ·04770, while my $\phi^2 = 2{\cdot}25$, giving as Coefficient of Mean Square Contingency $C_2 = {\cdot}83205$. We shall find that this drop down of the ψ^2 as we get lower and lower probabilities is the influence of the p_{ij} term in the numerator, and that, if we reach continuous variates, the influence of the approach of $\delta x \delta y$ to $dx dy$ and zero would be more and more conspicuous.

(7) Table VII shows a still greater increase in the scatter. We should anticipate lower values of ψ^2 and C_2, but still there is a very considerable degree of relationship between the two variates. We find $\psi^2 = {\cdot}01916$ and $\phi^2 = 1{\cdot}5$, giving a value ·77460 for C_2.

TABLE VII.

	·	·	·	·	·	·	·	·	·	·5	·5
	·	·	·	·	·	·	·	·	·5	·5	1
	·	·	·	·	·	·	·	·5	·5	·5	1·5
	·	·	·	·	·	·	·5	·5	·5	·5	2
	·	·	·	·	·	·5	·5	·5	·5	·5	2·5
	·	·	·	·	·5	·5	·5	·5	·5	·5	3
	·	·	·	·5	·5	·5	·5	·5	·5	·	3
	·	·	·5	·5	·5	·5	·5	·5	·	·	3
	·	·5	·5	·5	·5	·5	·5	·	·	·	3
	·5	·5	·5	·5	·5	·5	·	·	·	·	3
	·5	·5	·5	·5	·5	·	·	·	·	·	2·5
	·5	·5	·5	·5	·	·	·	·	·	·	2
	·5	·5	·5	·	·	·	·	·	·	·	1·5
	·5	·5	·	·	·	·	·	·	·	·	1
	·5	·	·	·	·	·	·	·	·	·	·5
	3	3	3	3	3	3	3	3	3	3	30

It is clear that the two methods differ completely in the positions in their respective scales, both running from 0 to 1, which are taken by the same table. It certainly seems as if the distributions dealt with in Tables I to VII ought to show high relationships between the variates, but actually judged by ψ^2 they give in all cases, but Table III, remarkably low values. We

attribute this in the first place to the presence of the factor p_{ij} in the numerator, which as we have shown at the beginning of this note tends to make ψ^2 vanish when the number of probabilities is large and their individual values are small*, except in the case of perfect contingency, when the value of ψ^2 seems suddenly to spring to unity. I do not therefore think that the values of ψ^2 can be very satisfactorily distributed through the range 0 to 1, since tables which under ordinary circumstances would be assumed to mark a high degree of relationship are nearly all crowded into the first quarter of the scale. Beside the advantage of ϕ^2 in this respect, and its approach to a finite limit when we pass to continuous variates; the coefficient of mean square contingency for an increasing number of cells passes into the correlation coefficient, on the hypothesis of normal distribution, and may even before that stage be fairly satisfactorily corrected for grouping.

* One other case was worked out with an extremely small scatter in the arrays, but with 2000 values p_{ij} differing from zero. Its form is indicated in Table VIII in the form of probabilities. The reader has to imagine this a contingency table of 1,001,000 cells with a narrow diagonal strip of only

TABLE VIII.

·	·	·	·	·	·	·	·	·	$\frac{1}{2000}$	$\frac{1}{2000}$
·	·	·	·	·	·	·	·	$\frac{1}{2000}$	$\frac{1}{2000}$	$\frac{1}{1000}$
·	·	·	·	·	·	·	$\frac{1}{2000}$	$\frac{1}{2000}$	·	$\frac{1}{1000}$
·	·	·	999 such rows				·	·	·	·
·	·	·	of $\frac{1}{2000}$, $\frac{1}{2000}$				·	·	·	·
·	$\frac{1}{2000}$	$\frac{1}{2000}$	·	·	·	·	·	·	·	$\frac{1}{1000}$
$\frac{1}{2000}$	$\frac{1}{2000}$	·	·	·	·	·	·	·	·	$\frac{1}{1000}$
$\frac{1}{2000}$	·	·	·	·	·	·	·	·	·	$\frac{1}{2000}$
$\frac{1}{1000}$	$\frac{1}{1000}$	$\frac{1}{1000}$	·	·	·	·	$\frac{1}{1000}$	$\frac{1}{1000}$	$\frac{1}{1000}$	1

2000 cells occupied, and think of this in terms of linear correlation! The correlation would be almost perfect. Steffensen's ψ^2 for this table is ·24975, and C_2=·99900, an almost perfect relationship.

(iii) Spearman's General Factor Again.

By BURTON H. CAMP.

In a recent paper† Piaggio has considered again questions having to do with the uniqueness of Spearman's general factor g, and has held that this uniqueness does not follow from the usual hypotheses and has produced further illustrations of cases where "almost uniqueness," as I called it in my *Biometrika* paper‡, does exist. Spearman's reply, adjoined to Piaggio's, apparently admits Piaggio's mathematics, and parenthetically mine, but asserts that our objections are not wholly relevant. It is tedious to drag out a controversy after the scientific evidence is once agreed on and published where everyone may examine it for himself, but I cannot but ask a brief space in order to point out as tersely as possible and without adjacent forbidding-looking mathematics two quite definite errors which it seems to me Spearman has made in his reply.

First, he says that the *kind* of indeterminateness that g has is not of the same sort as x, y, and z have in the equation of the plane,

$$Ax+By+Cz=0.$$

† *British Journal of Psychology*, Vol. xxiv. July 1933, pp. 88—105.

‡ Vol. xxiv. November 1932, pp. 418—427.

I think, on the contrary, that both Piaggio (p. 96, last line) and I (p. 424, Section 5) have shown explicitly that the indeterminateness of g is of exactly this type. For we have shown that the indeterminate factor (ki in Piaggio's notation, t in mine) cannot be determinate because it need satisfy only $K+1$ equations and it is itself a set of N unknowns, and N is greater than $K+1$. Similarly, Spearman's x, y, z cannot be determinate because they need to satisfy only one equation and they are themselves a set of 3 unknowns, and 3 is greater than 1. It seems to me impossible to recognise a difference in *type* between these two cases. It *might* be asserted that if N, the number of individuals tested, is small, the *range* of indeterminateness is limited by the numerical possibilities, because the standard deviation of ki is fixed, but these limits are exceedingly broad, and they increase indefinitely as N is increased*. Therefore such a claim could have no meaning from the psychological standpoint, for it cannot be true that the question whether individual Smith possesses a certain mental quality depends in some mathematical or occult way on how many Joneses are dragged in to be tested with him.

Secondly, Spearman says that the arbitrary term (ki) in the formula for g represents the "inexactitude" of one's *determination* of g, not a multiple nature in g itself. The error that he appears to me to be making here cannot be described so simply, but, I think, it is very important, especially as it is also implicit in the remarks of several other authors. It has been shown that, for the given set of $K-1$ tests and N individuals, there exists not one set of Ng's but an arbitrarily large group of such sets. Spearman appears to think that, because the variability of g *within each of these sets* is limited, it follows that the variability of a g from set to set for each fixed individual is also limited. This is not true, except in those extreme cases when $\sigma_{ki}=0$, exactly. Unfortunately Piaggio's language also encourages one to fall into this error, for in discussing his illustrative examples he implies that g can be shown to be practically unique if σ_{ki} for each set is negligibly small.

But it is easy to select a group of sets of ki's in which the variability within each set is as small as he indicates but for which the variability from set to set for any individual chosen arbitrarily in advance is very, very large. It is clearly, however, this last variability which must be proved to be small if it is to be shown in this manner that each individual has a characteristic g.

The gist of the matter is really contained in the mathematician's notion of "order of choice." If, before looking at Smith's scores on the tests, one may choose a number at random (subject only to the broad limitations mentioned before), and can then demonstrate that this number can be assigned as Smith's g, as well as any other number, and in perfect harmony with all the other hypotheses, then it is meaningless to assert that Smith *has* a g. In the numerical illustration at the close of my *Biometrika* paper, two such Smiths were exhibited (individuals 11 and 12).

It is not, of course, contended that a unique general factor does not, in truth, exist, but that its existence does not follow from Spearman's hypotheses. Moreover, it would seem inherently impossible that it could follow from his hypotheses, for into them he has introduced only group averages, of means, σ's, and r's. Out of them, therefore, whatever the mathematics, one could expect to obtain nothing more significant than characteristics of the group as a whole, and indeed this much, to repeat a statement in my other paper, has actually been established in the "almost unique" cases.

* *E.g.*, it is numerically impossible to choose 8 numbers for which the total range of variation is greater than 4σ, or 50 numbers for which it is greater than 10σ.

(iv) Note on the Recurrence Formulae for the Moments of the Point Binomial.

By A. A. KRISHNASWAMI AYYANGAR, Mysore University, Mysore, S. India.

§ 1. It is the object of this note to indicate that the recurrence formulae for the moments of the point binomial $(p+q)^n$ about the mean nq can be most simply obtained by a little ingenious transformation of the term $(t-nq)^s p^{n-t}q^t \binom{n}{t}$ in the series for μ_s the sth moment. This transformation explains why the recurrence formulae for incomplete moments are as easy to derive as those for complete moments, which are particular cases of the former.

We may write

$$(t-nq)^s p^{n-t}q^t\binom{n}{t}=nq\,(t-1-\overline{n-1}\,.\,q+p)^{s-1}p^{n-t}q^{t-1}\binom{n-1}{t-1}-nq\,(t-nq)^{s-1}p^{n-t}q^t\binom{n}{t}$$

[since $p+q=1$] .....(1).

Let $$\nu_s{}^\rho=\sum_{t=\rho}^{t=n}(t-nq)^s p^{n-t}q^t\binom{n}{t}.$$

Then

$$\frac{1}{nq}\,.\,\nu_s{}^\rho=\sum_{t=\rho}^{t=n}(t-1-\overline{n-1}\,.\,q+p)^{s-1}p^{n-t}q^{t-1}\binom{n-1}{t-1}-\sum_{t=\rho}^{t=n}(t-nq)^{s-1}p^{n-t}q^t\binom{n}{t}$$

$$=\sum_{t'=\rho-1}^{t'=n-1}(t'-\overline{n-1}\,.\,q+p)^{s-1}p^{n-t'-1}q^{t'}\binom{n-1}{t'}-\sum_{t=\rho}^{t=n}(t-nq)^{s-1}p^{n-t}q^t\binom{n}{t},$$

i.e. $$\nu_s{}^\rho=\{(E+p)^{s-1}\nu_0{}'^{\rho-1}-\nu_{s-1}{}^\rho\}\,nq \quad\text{.....(2),}$$

where $E(\nu_s{}'^{\rho-1})=\nu_{s+1}{}^{\rho-1}$ and $\nu_s{}'^{\rho-1}$ is the sth incomplete moment defined for the binomial $(p+q)^{n-1}$. Putting $\rho=0$, we get the recurrence formula for complete moments

$$\mu_s=-nq\,\mu_{s-1}+nq\,(E+p)^{s-1}\mu_0{}' \quad\text{.....(3).}$$

The formulae (2) and (3) are believed to be new.

Application:

For brevity we write ν_s and $\nu_s{}''$ for $\nu_s{}^\rho$ and $\nu_s{}'^{\rho-1}$. Putting $s=1$ in (2), we have

$$\nu_1=(\nu_0{}''-\nu_0)\,nq, \text{ i.e. } \nu_0{}''=\nu_0+\frac{\nu_1}{nq} \quad\text{.....(4).}$$

But

$$\nu_0{}''=\sum_{t=\rho-1}^{t=n-1}p^{n-1-t}q^t\binom{n-1}{t}$$

$$=\sum_{t=\rho-1}^{t=n-2}\left[p^{n-1-t}q^{t+1}\binom{n}{t+1}+p^{n-t}q^t\binom{n-1}{t}-p^{n-\overline{t+1}}q^{t+1}\binom{n-1}{t+1}\right]+q^n\binom{n}{n}+pq^{n-1}\binom{n-1}{n-1}$$

since $\binom{n-1}{t}=\binom{n-1}{t}p+\binom{n}{t+1}q-\binom{n-1}{t+1}q$

$$=\sum_{t'=\rho}^{t'=n}p^{n-t'}q^{t'}\binom{n}{t'}+\sum_{t=\rho-1}^{t=n-2}\left[p^{n-t}q^t\binom{n-1}{t}-p^{n-\overline{t+1}}q^{t+1}\binom{n-1}{t+1}\right]+pq^{n-1}\binom{n-1}{n-1},$$

i.e. $$\nu_0{}''=\nu_0+p^{n-\rho+1}q^{\rho-1}\binom{n-1}{\rho-1} \quad\text{.....(5).}$$

Comparing (4) and (5), we get

$$\nu_1=np^{n-\rho+1}q^\rho\binom{n-1}{\rho-1}=\rho p^{n-\rho+1}q^\rho\binom{n}{\rho}^* \quad\text{.....(6).}$$

* This result, for which we have given here a simple proof, is due to Ragnar Frisch. See *Scandinavisk Actuarietidskrift* (1924), No. 3, p. 161, quoted in *Biometrika*, Vol. XVII. (1925) p. 170.

Hence $$\nu_1'' = \nu_1(\rho - 1)/nq \quad \ldots\ldots(7).$$

Similarly by setting $s = 2, 3, \ldots$ in (2) we calculate in succession

$$\left.\begin{aligned}
\nu_2 &= \nu_1(\rho - \overline{n+1}\,.\,q) + \nu_0 npq \,* \\
\nu_2'' &= \nu_1\left\{p - \frac{p}{n} + \frac{(\rho-1)^2}{nq} - \rho + 1\right\} + \nu_0 pq(n-1) \\
\nu_3 &= \nu_1\{(\rho - \overline{n+1}\,.\,q)^2 + pq(2n-1)\} + \nu_0 npq(p-q)\,* \\
\nu_3'' &= \frac{\nu_1}{nq}\{(\rho-1)[(\rho-1-nq)^2 + pq\,.\,\overline{2n-3}] + pq(p-q)(n-1)\} \\
&\qquad + \nu_0(n-1)pq(p-q) \\
\nu_4 &= \nu_1[(\rho - \overline{n+1}\,.\,q)^3 + 3pq(n-1)(\rho - \overline{n+1}\,.\,q) + pq(p-q)(3n-1)] \\
&\qquad + \nu_0[npq(1 + 3pq\,.\,\overline{n-2})] \\
&\text{and so on}
\end{aligned}\right\} \quad \ldots\ldots(8).$$

In the above expressions, the coefficients of ν_0 are the complete moment coefficients and follow the recurrence-relation (3); and it must be possible to discover a recurrence relation for the coefficients of ν_1 also (see the following section).

§ 2. We will now give a simple proof of Pearson's recurrence formula †, which enables us also to write down immediately the corresponding recurrence formula for incomplete moments given by Ragnar Frisch‡.

We transform $(t-nq)^{s-1}p^{n-t}q^t\binom{n}{t}$ in two different ways:

$$(1)\quad pt(t-nq)^{s-1}p^{n-t}q^t\binom{n}{t} = p(t-nq)^s p^{n-t}q^t\binom{n}{t} + nqp(t-nq)^{s-1}p^{n-t}q^t\binom{n}{t},$$

$$(2)\quad pt(t-nq)^{s-1}p^{n-t}q^t\binom{n}{t} = (t-nq)^{s-1}p^{n-t+1}q^{t-1}\binom{n}{t-1}(n-t+1)q$$

$$= (t-nq)^{s-1}p^{n-t+1}q^{t-1}\binom{n}{t-1}(nqp - q\,.\,\overline{t-1-nq})$$

since $p+q=1$.

$$\therefore\ p\{(t-nq)^s + nq(t-nq)^{s-1}\}p^{n-t}q^t\binom{n}{t} = (t-nq)^{s-1}p^{n-\overline{t-1}}q^{t-1}\binom{n}{t-1}(nqp - q\,.\,\overline{t-1-nq}) \quad \ldots\ldots(9).$$

$$\therefore\ p\sum_{t=\rho}^{t=n}\{(t-nq)^s + nq(t-nq)^{s-1}\}p^{n-t}q^t\binom{n}{t}$$

$$= \sum_{t'=\rho-1}^{t'=n-1}\left\{(t'-nq+1)^{s-1}p^{n-t'}q^{t'}\binom{n}{t'}\right\}(nqp - q\,.\,\overline{t'-nq})$$

$$= \sum_{t'=\rho}^{t'=n}\left\{(t'-nq+1)^{s-1}p^{n-t'}q^{t'}\binom{n}{t'}\right\}(nqp - q\,.\,\overline{t'-nq})$$

$$+ (\rho-nq)^{s-1}p^{n-\rho+1}q^{\rho-1}\binom{n}{\rho-1}(nqp - q\,.\,\overline{\rho-1-nq})$$

$$= \sum_{t'=\rho}^{t'=n}\left\{(t'-nq+1)^{s-1}p^{n-t'}q^{t'}\binom{n}{t'}\right\}(nqp - q\,.\,\overline{t'-nq}) + (\rho-nq)^{s-1}\rho p^{n-\rho+1}q^{\rho}\binom{n}{\rho},$$

i.e. $$p(E^s + nqE^{s-1})\nu_0 = (E+1)^{s-1}(nqp\nu_0 - q\nu_1) + \nu_1(\rho - nq)^{s-1},$$

i.e. $$\{(1+E)^{s-1} - E^{s-1}\}(nqp\nu_0 - q\nu_1) = \nu_s - \nu_1(\rho - nq)^{s-1} \quad \ldots\ldots(10).$$

If we put here $\rho = 0$, then $\nu_1 = 0$ and we get Pearson's formula.

* These results were given previously by Ragnar Frisch.

† *Biometrika* (1924), Vol. XVI. p. 160.

‡ *Ibid.* (1925), Vol. XVII. p. 171.

From (8) we see that ν_s is of the form $\nu_1\lambda_s+\nu_0\mu_s$, where $\lambda_0=0$ and $\lambda_1=1$. Using this expression in (10) and equating the coefficients of ν_1, we get the recurrence relation for λ_s, viz.

$$\{(1+E)^{s-1}-E^{s-1}\}(nqp\lambda_0-q\lambda_1)=\lambda_s-(\rho-nq)^{s-1} \quad \ldots\ldots\ldots\ldots\ldots\ldots(11),$$

which was not obvious from the other formula (2).

(v) A Note on the Incomplete Moments of the Hypergeometrical Series.

By A. A. KRISHNASWAMI AYYANGAR, Maharajah's College, Mysore, South India.

1. The aim of this note is to indicate an extremely simple and elementary method, applicable alike to complete and to incomplete moments, of deriving the recurrence formula* due to Karl Pearson for the moments of the hypergeometrical series defined by

$$\frac{pn\,(pn-1)\ldots(pn-r+1)}{n\,(n-1)\ldots(n-r+1)}\left[1+\frac{r}{1}\cdot\frac{qn}{pn-r+1}\,x+\frac{r\,(r-1)}{1\,.\,2}\cdot\frac{qn\,(qn-1)}{(pn-r+1)\,(pn-r+2)}\,x^2+\ldots\right].$$

Here the coefficient of x^t, which we may denote by $\phi(t)$ for brevity, is the probability of drawing t individuals without some mark in a sample of r individuals selected at random from a population n containing pn marked and qn unmarked individuals. The mean of the samples is evidently rq.

The incomplete moment of the sth order about the mean is defined by

$$\nu_s=\sum_{t=\rho}^{r}(t-rq)^s\,\phi(t).$$

2. The following identities are self-evident:

(i) $$t\,(pn-r+t)=n\,(t-rq)+(r-t)\,(qn-t);$$

(ii) $$(r-t)\,(qn-t)=(t-rq)^2-c_1\,(t-rq)+c_2,$$

where $$c_1=nq+r\,(p-q) \quad \text{and} \quad c_2=rpq\,(n-r);$$

(iii) $$t\,(pn-r+t)\,\phi(t)=(r-\overline{t-1})\,(qn-\overline{t-1})\,\phi(t-1).$$

Multiplying both sides of (iii) by $(t-rq)^{s-1}$ and summing for values of t from ρ to r, we have using (i) and (ii),

$$\begin{aligned}
&\sum_{t=\rho}^{r}\phi(t)\,(t-rq)^{s-1}\{n\,(t-rq)+(t-rq)^2-c_1\,(t-rq)+c_2\}\\
&\quad=\sum_{t=\rho}^{r}\phi(t-1)\,(t-rq)^{s-1}\,(r-\overline{t-1})\,(qn-\overline{t-1})\\
&\quad=\sum_{t'=\rho-1}^{r-1}\phi(t')\,(t'-rq+1)^{s-1}\,(r-t')\,(qn-t'), \quad \text{where } t'=t-1,\\
&\quad=\sum_{t'=\rho}^{r}\phi(t')\,(t'-rq+1)^{s-1}\{(t'-rq)^2-c_1\,(t'-rq)+c_2\}\\
&\qquad+\phi(\rho-1)\,.\,(\rho-rq)^{s-1}\,(r-\rho+1)\,(qn-\rho+1).
\end{aligned}$$

Introducing the operator E to denote the operation

$$E(\nu_s)=\nu_{s+1},$$

we write the above result in the form

$$n\nu_s+E^{s-1}(\nu_2-c_1\nu_1+c_2\nu_0)=(1+E)^{s-1}(\nu_2-c_1\nu_1+c_2\nu_0)+(\rho-rq)^{s-1}\,\rho\,(pn-r+\rho)\,\phi(\rho),$$

i.e. $$\{(1+E)^{s-1}-E^{s-1}\}(\nu_2-c_1\nu_1+c_2\nu_0)=n\nu_s-(\rho-rq)^{s-1}\,\rho\,(pn-r+\rho)\,\phi(\rho).$$

* *Biometrika*, Vol. XVI. (1924) p. 159.

Put $s=1, \quad n\nu_1=\rho\,(pn-r+\rho)\,\phi\,(\rho),$

i.e. $$\nu_1=\frac{\rho\,(pn-r+\rho)}{n}\,\phi(\rho)^* \qquad \text{.....(1).}$$

Hence, we may write the recurrence formula for incomplete moments in the form

$$\{(1+E)^{s-1}-E^{s-1}\}\,(\nu_2-c_1\nu_1+c_2\nu_0)=n\nu_s-n\,(\rho-rq)^{s-1}\nu_1 \qquad \text{.....(2).}$$

This shows that we may put

$$\nu_s=\mu_s\nu_0+\lambda_s\nu_1,$$

where $$\lambda_0=\mu_1=0, \quad \lambda_1=\mu_0=1\dagger,$$

in (2) and equate the coefficients of ν_1 and ν_0 and obtain recurrence formulae separately for λ_s and μ_s.

In fact, the recurrence formula thus obtained for μ_s is the same as Pearson's for complete moments, while that for λ_s is

$$\{(1+E)^{s-1}-E^{s-1}\}\,(\lambda_2-c_1\lambda_1+c_2\lambda_0)=n\,\{\lambda_s-(\rho-rq)^{s-1}\} \qquad \text{.....(3).}$$

Put $$s=2, \quad \lambda_2-c_1=n\lambda_2-n\,(\rho-rq),$$

i.e. $$\lambda_2=\frac{n(\rho-q\,.\,\overline{r+1})-r(p-q)}{n-1} \qquad \text{.....(4).}$$

Put $s=3, \quad (1+2E)\,(\lambda_2-c_1\lambda_1+c_2\lambda_0)=2\lambda_3+(1-2c_1)\,\lambda_2+(2c_2-c_1)=n\lambda_3-n\,(\rho-rq)^2.$

Therefore $$(n-2)\,\lambda_3=n\,(\rho-rq)^2+(1-2c_1)\,\frac{n\,(\rho-rq)-c_1}{n-1}+2c_2-c_1,$$

i.e. $$\lambda_3=\frac{1}{(n-1)\,(n-2)}\,[n^2\,\{(\rho-\overline{r+1}q)^2+pq\,(2r-1)\}$$
$$-n\,\{(\rho-\overline{r+1}\,.\,q)^2+(\rho-\overline{r+1}\,.\,q)\,(2r-1)\,(p-q)$$
$$+pq\,(2r^2-2r-1)+r\}+2r^2\,(1-3pq)] \qquad \text{.....(5).}$$

When $\rho=0$, it follows from (1) that $\nu_1=0$ and obviously $\nu_0=1$, so that $\nu_s=\mu_s$, which satisfies Pearson's recurrence formula.

(vi) On Simometers and their Handling.

By K. PEARSON.

The original purpose of the simometer was to measure the flatness of the nasal bridge by the ratio of its subtense to the chord. For this purpose we may take either the simotic or the rhinal index, or indeed both. But for the purposes of this note we are not concerned with what terminals may be chosen for our fundamental chord; we are concerned only with what is the definition of our "subtense." Probably to most mathematicians the subtense is the portion of the perpendicular bisector of the chord intercepted between the chord and the arc. It was from this conception of a subtense that Newton proceeded to his definition of curvature. If the arc while still an arc of continuous curvature be not symmetrical about the chordal bisector but skew to it, then the line from the point of chordal bisection to the point of contact of the tangent parallel to the chord has been considered—as in the cases of parabola and ellipse—as the subtense.

Now in the case of the nasal bridge when the two sides of the nose are not mirror images the one of the other, the length measured as "subtense," either by the simometer of Mérejkowsky or by the co-ordinate calipers of Woo (see pp. 197—98 of this issue), is not the mathematical subtense as just defined. What, however, is of more importance, the two instruments do not, or should not, give the same result, and even the same instrument will give different results according

* This is a generalisation of Frisch's result given in *Biometrika*, Vol. XVII. (1925) p. 170 (18).

† These values have to be used only after the operator E has had its effect.

to the manner in which its use is prescribed. Both Mérejkowsky's and Woo's simometers possess two degrees of freedom for the measuring tip of the mid-arm. We will consider their geometrical aspects separately.

Let *ab* be the chord and *afdb* the arc of which the subtense is to be measured. Then when the tips of the curved outer arms of Mérejkowsky's simometer are placed on *a* and *b* there is a fixed point *g* of the instrument at equal distances from *a* and *b*. Through *g* a line passes (actually a rod), the end of which, *d* (the tip of the rod or mid-leg), can be adjusted to any point of the area *agb* by a turn round *g* and a forward motion of the tip *d* produced by a screw reading to $\frac{1}{10}$ mm. If *gd* be taken to *bisect* the angle *agb*, then Mérejkowsky's will enable us to read off the subtense *cd*, that is the length of the line through the bisection of *ab*, perpendicular to *ab*, and reaching to the curve. This requires two readings of the instrument, first to measure *gd* directly, and second to measure *gc*; the latter can only be obtained by retaining the outer legs at the distance *ab*, removing the simometer from the nasal bridge and, without rotating *gd*, bringing *d* down to *c* so that the lips of the three tips of the legs all touch a plane.

If the tip *d* be not kept to the perpendicular bisector *cg*, but directed to some point *h* on the arc, then Mérejkowsky's simometer measures *hi*, which has no pretence to be a subtense, and may be greater than the maximum distance *fe* of the arc from the chord. The true maximum subtense will only be given by this instrument when the nose is symmetrical and care is taken that the mid-leg bisects the angle *agb*. This simometer provides no means of ascertaining directly the point *f* of the arc most remote from *ab*, and if the tip of the mid-leg be brought to *f* the subtense *fj*

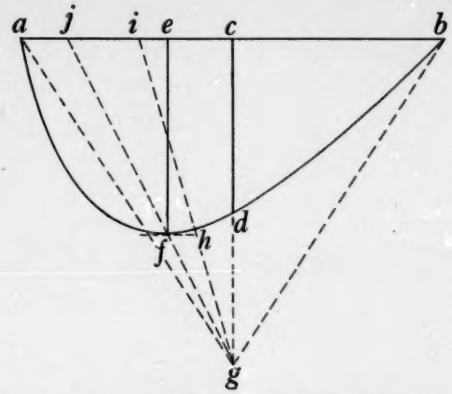


Fig. 1. *ag*=*gb*, *cg* is a perpendicular bisector of *ab*, *f* is highest point of arc *afb*, and *h* is an arbitrary point between *f* and *d*.

might in a very skew nasal bridge be such that *j* did not even fall inside *ab*! Further, we have seen that *f* cannot be ascertained from this instrument. What the simometer can achieve is the ascertainment of the minimum distance from *ab* of the trace of the nasal bridge on the plane perpendicular to and bisecting the chord *ab*.

We now turn to Dr Woo's simometer or his co-ordinate calipers*.

It has three legs perpendicular to a bar, of which the second (*gf*) and third (*mb*) can be shifted along the bar and the second or mid-leg moved also perpendicular to the bar: all three legs and the bar may be discussed geometrically as coplanar lines. Fixing the external *la* and *mb* with their

* This name does not seem very well chosen.

tips at a and b, the mid-leg, gnf, has two degrees of freedom, and there is no difficulty in bringing it into the position f which enables the maximum subtense of the chord ab to be read off. By turning the instrument round ab and reading off a number of maximum subtenses like ef, we can find their minimum, which is probably the quantity of most value to the anthropologist. But unfortunately Dr Woo does not appear to have used his simometer in this manner, any more than the later workers in the Biometric Laboratory have used Mérejkowsky's simometer with the mid-leg bisecting the angle between the chords of the external legs. The procedure in both cases has been apparently to draw on the nasal bridge a line supposed to represent the nasal bridge,

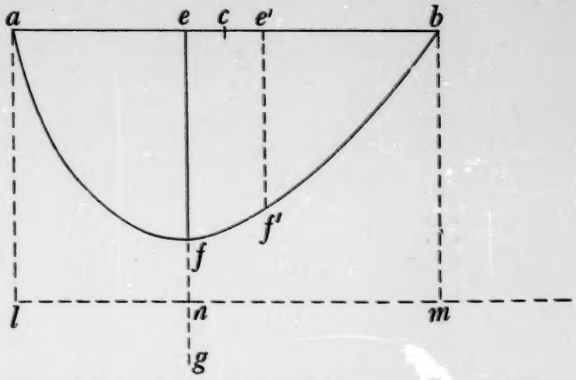


Fig. 2. c is mid-point of ab, $al=bm$, f is highest point of asymmetrical arc afb, $efng$ is perpendicular to ab, f' is an arbitrary point, $f'e'$ is perpendicular to ab, g is a point on efn produced, and lm is produced beyond m to some distance.

and this seems to have been done without any reference to the chord ab. Its obvious purpose is to get rid of one of the two degrees of freedom, which both instruments possess when working in a definite plane. But if we pay no regard to the chord ab, f may well be at f' instead of the point at maximum distance from ab, and accordingly Dr Woo's simometer would measure $f'e'$, not fe. Even with a nasal bridge which without regard to sutures had perfect mirror symmetry, both simometers would fail to give the maximum subtense if a and b were not images one of the other in the mirror plane of that nasal bridge, and we aimed our mid-leg tip at the point where the arc met the mirror plane.

To sum up: Both instruments seem to have been used in a manner which, at any rate *theoretically*, invalidates their purpose, if that purpose be to find the minimum subtense (perpendicular to ab) of all the maximum subtenses of sections of the nasal bridge by planes through ab*.

Woo's instrument could give the required result, but with more labour than is necessary to obtain a doubtful result from an arbitrary median nasal ridge. Mérejkowsky's simometer could never give in asymmetrical nasal bridges (including the asymmetry of a and b as a part of the nasal) the minimum of the maximum subtenses, but if we start by defining the median nasal ridge to be the trace on the nasal bridge of the plane perpendicularly bisecting ab, it will give the minimum subtense in this plane by simply causing the mid-leg to bisect the angle at g. No system which does not define the median nasal ridge with reference to ab can it seems to me be really satisfactory.

* Optically we may suppose the eye placed at a great or infinite distance along the line ab produced. Such an eye would see the nasal bridge as a plane curve (although all the points of the curve might not be coplanar), and the point on this curve nearest to the line ab would give the required subtense.

The optical conception could be carried out excellently with the cranial co-ordinatograph*. Place the skull by aid of it with *ab* perpendicular to the drawing board, and trace with the knife edge projector the contour of the nasal bridge; the minimum distance from the plan of the line *ab* to this contour would be the subtense sought. This frees us from any arbitrary choice of a median ridge for the nasal bridge or from the supposition that the maximum subtenses of all the sections through the chord lie in one plane, the trace of which can be determined by personal appreciation.

I am inclined to believe that the simometer of Mérejkowsky, with its choice of the plane of bisection of the chord as plane for measuring the subtense, is at least as accurate as Dr Woo's calipers applied to an arbitrary median ridge which pays no attention to the chord. What the latter method would achieve if used as a three-degree of freedom instrument to determine the minimum of the maximum subtenses it is difficult to say without experimenting. What theoretically seems to me erroneous is the determination of a nasal ridge without direct reference to the chord *ab*. That the two instruments do not lead to widely, although they do to significantly different results is consoling to the craniologist, if not to the theorist. There are certainly some erratic differences in the tables on p. 200, which require clearing up.

It may not be out of place to recall here a method I introduced in 1911 for determining the flatness of the nasal bridge. A fine celluloid tape or even a thread may be used to measure the shortest arcual or geodesic distance from the terminals of the chord *ab* across the nasal bridge. A steel tape is unsuitable. The chord *ab* is also measured. A very suitable measure for the flatness of the nose is then

$$\beta = 100\,(\text{arc} - \text{chord})/\text{arc}.$$

This is adequate without any determination of the subtense.

But, if we desire to find the subtense, it is needful to choose a curve sufficiently elastic to represent the nasal bridge. The circle was no use, the parabola somewhat better, but finally, after drawing a good many nasal bridge sections, I concluded that the catenary gave the best result. If then

$$\alpha = 100\ \text{subtense}/\text{chord},$$

it is needful to compute α for each value of β. Tables to find α given β were calculated by H. E. Soper and Dr Julia Bell†.

It must be admitted at once that like Mérejkowsky's simometer this method starts with the assumption of a symmetrical nasal bridge. But if the nasal bridge be skew, the increased length of arc will to some extent compensate in an increased subtense, which compensation is lacking in the case of the simometer.

It would, I think, be of actual practical value, especially as the racial value of the simotic index is becoming more generally acknowledged, to take a series of, say, one hundred skulls, and measure the chord *ab*, the arc *afb* and the minimum of the maximum subtenses, using

(*a*) The cranial co-ordinatograph to give "the true value";

(*b*) Mérejkowsky's simometer with the mid-leg as bisector‡;

(*c*) Woo's simometer, with no attempt at personal appreciation of a median nasal ridge, and

(*d*) The catenary method.

Such an experiment would enable us, with due regard to accuracy of result, to the time and to the labour involved, to select the practically best method and to discard the others.

* See *Biometrika*, Vol. xxv. pp. 217—253.

† Published originally in *Biometrika*, Vol. viii. pp. 316, 318 and Vol. ix. pp. 401—2; reissued *Tables for Statisticians and Biometricians*, Part i. pp. lvi—lvii, 32—64.

‡ It would not be difficult to devise a mechanically less crude instrument of the Mérejkowsky type than the present one.